

UIT-R

Sector de Radiocomunicaciones de la UIT

Recomendación UIT-R BT.1867
(03/2010)

Técnicas de medición objetiva de la calidad visual percibida para aplicaciones de radiodifusión que utilizan la televisión de definición reducida 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, 2010

© UIT 2010

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.1867

Técnicas de medición objetiva de la calidad visual percibida para aplicaciones de radiodifusión que utilizan la televisión de definición reducida* en presencia de una referencia de anchura de banda reducida**

(2010)

Cometido

Esta Recomendación especifica métodos para la estimación de la calidad visual percibida para aplicaciones de radiodifusión que utilizan televisión de definición reducida (TVDR) cuando la señal disponible es una señal de referencia reducida (RR), por ejemplo, a través de un canal de datos auxiliar, marca de agua, metadatos u otros.

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 de enorme interés para la industria
- b) que la Recomendación UIT-R BT.1683 describe 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;
- c) que la Recomendación UIT-R BT.1833 describe sistemas multimedia para la radiodifusión de multimedia y aplicaciones de datos para recepción móvil mediante receptores de bolsillo;
- d) que la televisión de definición reducida (TVDR) está empezando a ser ampliamente utilizada en la radiodifusión de multimedia y aplicaciones de datos para recepción móvil;
- e) que la Recomendación UIT- J.246¹ especifica técnicas de medición objetiva de la calidad de vídeo perceptual para aplicaciones de televisión de definición reducida en presencia de una referencia de anchura de banda reducida;
- f) que la medición objetiva de la calidad de vídeo percibida puede resultar útil como complemento de los métodos de evaluación subjetiva,

* La televisión de definición reducida (TVDR) hace referencia a resoluciones de vídeo con un número de pixels menor que los definidos en la Recomendación UIT-R BT.601. Está en estudio una Recomendación UIT-R pertinente sobre TVDR.

** El método de medida con referencia reducida para la medición objetiva de la calidad percibida de vídeo, evalúa la calidad de funcionamiento de los sistemas mediante comparaciones entre características extraídas de la señal de vídeo no distorsionada a la entrada del sistema y la señal degradada a la salida del sistema.

¹ La Recomendación UIT-T J.246 está disponible en <<http://www.itu.int/rec/T-REC-J.246-200808-P/en>>.

reconociendo

a) que la utilización de TVDR está destinada principalmente a la visualización en pantallas pequeñas, tales como las disponibles en receptores portátiles y móviles,

recomienda

1 que las directrices, cometido y limitaciones del Anexo 1 se utilicen en la aplicación de modelos de medición objetiva de vídeo identificados en el *recomienda 2*;

2 que el modelo de medición objetiva de la calidad visual percibida del Anexo 2 se utilice en aplicaciones de radiodifusión que utilicen LDTV cuando está disponible una señal de referencia reducida, tal como se describe en el Anexo 2

Anexo 1

1 Introducción

En esta Recomendación se especifican métodos para la estimación de la calidad de vídeo percibida en aplicaciones de radiodifusión que utilizan TVDR cuando está disponible una señal de referencia reducida.

El método de medición de referencia reducida puede utilizarse cuando las características extraídas de la señal de vídeo de referencia están disponibles en el punto de medida, como ocurre en mediciones realizadas sobre equipos individuales, en una cadena en el laboratorio o en un entorno cerrado. Los métodos de estimación están basados en el procesado de vídeo con resolución VGA, CIF y QCIF.

El material de prueba de validación utilizado presentaba diversas degradaciones de codificación y errores de transmisión (por ejemplo, bit erróneos, paquetes descartados). Si existen señales de vídeo con distorsiones de codificación, el codificador puede utilizar varios métodos de compresión (por ejemplo, MPEG-2, H.264, etc.). Los modelos de esta Recomendación pueden utilizarse para la supervisión de la calidad en redes a fin de garantizar su disponibilidad operacional. Los efectos visuales de las degradaciones pueden ser espaciales y temporales (por ejemplo, repetición de cuadros, omisión de cuadros, reducción de la velocidad de cuadros). Los modelos de esta Recomendación también pueden utilizarse para pruebas de laboratorio de sistemas de vídeo. Cuando se utilizan para comparar diferentes sistemas de vídeo, es recomendable utilizar un método cuantitativo (como el de la Recomendación UIT-T J.149) para determinar la exactitud del modelo en dicho contexto en particular. Esta Recomendación se considera apropiada para servicios entregados a una velocidad de 4 Mbit/s o menor, típica de receptores móviles. Las pruebas de validación de cada resolución se realizaron en las condiciones siguientes:

- QCIF (*quarter common intermediate format*) (formato de un cuarto del formato intermedio común (176 × 144 píxeles)): 16 kbit/s a 320 kbit/s.
- CIF (*common intermediate format*) formato intermedio común (352 × 288 píxeles): 64 kbit/s – 2 Mbit/s.
- VGA (*video graphics array*) matriz de gráficos de vídeo (640 × 480 píxeles): 128 kbit/s – 6 Mbit/s .

CUADRO 1

Factores utilizados en la evaluación de modelos

Factores de prueba
Errores de transmisión con pérdida de paquetes.
Resolución de vídeo QCIF, CIF y VGA.
Velocidades binarias de vídeo: – QCIF: 16 kbit/s a 320 kbit/s – CIF: 64 kbit/s – 2 Mbit/s – VGA: 128 kbit/s – 4 Mbit/s.
Errores temporales (pausas con omisión de imágenes) de 2 s como máximo.
Velocidades de cuadros de vídeo de 5 a 30 cuadros/s.
Métodos de codificación
H.264/AVC (MPEG-4 Parte 10), MPEG-4 Parte 2, y otros tres métodos de codificación privados. (Véase Nota 1).
Aplicaciones
Supervisión de la calidad en tiempo real y en funcionamiento normal.
Supervisión de la calidad en el destino distante cuando hay disponibles canales paralelos para características extraídas de secuencias de vídeo de la fuente.
Medición de la calidad para la supervisión de un sistema de almacenamiento o de transmisión que utilice técnicas de compresión y descompresión de vídeo, ya sea en un solo paso o mediante concatenación de dichas técnicas.
Prueba en laboratorio de sistemas de vídeo.

NOTA 1 – Las pruebas de validación de modelos incluyeron secuencias de vídeo codificadas utilizando 15 códecs de vídeo diferentes. Los cinco códecs que figuran en el Cuadro 1 se utilizaron para la codificación de secuencias de prueba, y cualquiera de los modelos recomendados puede considerarse adecuado para la evaluación de dichos códecs. Además de los cinco códecs mencionados, una pequeña parte de las secuencias de prueba se crearon utilizando los códecs siguientes: H.261, H.263, H.263+², JPEG-2000, MPEG-1, MPEG-2, H.264 SVC, y otros sistemas privados. Nótese que algunos de dichos códecs se utilizaron solamente para resoluciones CIF y QCIF pues es previsible que su uso real sea principalmente con dichas resoluciones.

Antes de la aplicación de un modelo a secuencias codificadas utilizando uno de dichos códecs, es recomendable que el usuario examine cuidadosamente sus prestaciones predictivas para determinar si el modelo alcanza unas características de funcionamiento predictivas adecuadas.

2 Aplicaciones

Las aplicaciones para los modelos de estimación descritos en esta Recomendación incluyen, aunque no exclusivamente, las siguientes:

- 1) evaluación, especificación y pruebas de aceptación de códecs, coherentes con una precisión limitada, tal como se describe más adelante;
- 2) supervisión de la calidad en tiempo real y en funcionamiento;

² H.263+ es una configuración particular de H.263 (1998).

- 3) supervisión distante de la calidad del destino cuando hay disponibles canales laterales que permiten disponer de las características extraídas de secuencias de vídeo de la fuente;
- 4) medición de la calidad para la supervisión de un sistema de almacenamiento o de transmisión que utilice técnicas de compresión y descompresión de vídeo, ya sea en un solo paso o mediante la concatenación de dichas técnicas;
- 5) Prueba en laboratorio de sistemas de vídeo.

3 Limitaciones

Los modelos de estimación descritos en esta Recomendación no deben utilizarse para sustituir completamente a las pruebas subjetivas. Los valores de la correlación entre dos pruebas subjetivas cuidadosamente diseñadas y realizadas (es decir, en dos laboratorios diferentes) está normalmente entre 0,95 y 0,98. Si este modelo se utiliza para comparar sistemas de vídeo (por ejemplo, la comparación de dos códecs), es recomendable utilizar un método cuantitativo (como el de la Recomendación UIT-T J.149) para determinar la precisión del modelo en dicho contexto en particular.

Los modelos incluidos en esta Recomendación se han validado por la medición de vídeos con congelaciones de cuadros de imagen de hasta 2 s.

Los modelos de esta Recomendación no se han validado para la medición de vídeo con un retardo creciente de forma persistente (por ejemplo, vídeos que no descarten tramas perdidas tras un congelamiento de cuadro).

Debe señalarse que en el caso de nuevas tecnologías de codificación y transmisión que introduzcan efectos secundarios no considerados en esta evaluación, los modelos objetivos pueden producir resultados erróneos. En ese caso, es necesaria una evaluación subjetiva.

4 Descripción de modelos

En el Anexo 2 se describen los modelos siguientes:

Modelo A (Anexo 2) – VQEG propuesto por la Universidad Yonsei, de la República de Corea.

Apéndice 1 al Anexo 1

Resultados del Grupo de Expertos en Calidad del Vídeo (VQEG)

Los estudios relativos a mediciones de la calidad de vídeo percibida son realizados por un grupo informal denominado VQEG (*Video Quality Experts Group*), que está integrado en las Comisiones de Estudio 9 y 12 del UIT-T y en la Comisión de Estudio 6 del UIT-R. En la recién finalizada Fase I de pruebas de Multimedia del VQEG se han evaluado las prestaciones de los algoritmos propuestos para la medición de calidad percibida de vídeo con referencia reducida para los formatos QCIF, CIF y VGA.

En base a la evidencia actualmente reconocida, el Sector de Radiocomunicaciones de la UIT recomienda el método siguiente:

Modelo A (Anexo 2) – VQEG propuesto por la Universidad Yonsei, de la República de Corea.

Los Cuadros 2, 3 y 4 proporcionan información de detalle sobre las prestaciones del modelo en las pruebas realizadas durante la Fase I de multimedios con VQEG.

CUADRO 2

Resolución VGA: prestaciones del modelo en las pruebas de la Fase I de multimedios con VQEG: valores medios en relación con 13 pruebas subjetivas

Estadística	Yonsei RR10k	Yonsei RR64k	Yonsei RR128k	PSNR ⁽¹⁾
Correlación	0,803	0,803	0,803	0,713
RMSE ⁽²⁾	0,599	0,599	0,598	0,714
Relación de afloramiento	0,556	0,553	0,552	0,615

⁽¹⁾ PSNR: relación señal a ruido de cresta.

⁽²⁾ RMSE: error cuadrático medio.

CUADRO 3

Resolución CIF: prestaciones del modelo en las pruebas de la Fase I de multimedios con VQEG: valores medios en relación con 14 pruebas subjetivas

Estadística	Yonsei RR10k	Yonsei RR64k	PSNR
Correlación	0,780	0,782	0,656
RMSE	0,593	0,590	0,720
Relación de afloramiento	0,519	0,511	0,632

CUADRO 4

Resolución QCIF: prestaciones del modelo en las pruebas de la Fase I de multimedios con VQEG: valores medios en relación con 14 pruebas subjetivas

Estadística	Yonsei RR1k	Yonsei RR10k	PSNR
Correlación	0,771	0,791	0,662
RMSE	0,604	0,578	0,721
Relación de afloramiento	0,505	0,486	0,596

Las correlaciones medias del análisis básico de los modelos de VGA con referencia reducida (RR) tuvieron todas el valor 0,80, con una PSNR de 0,71. Las correlaciones del modelo individual para algunos de los experimentos fueron de hasta 0,93. El RMSE medio de los modelos de VGA con RR

fueron todos de 0,60, con una PSNR de 0,71. La relación de afloramiento media para los modelos de VGA con RR osciló entre 0,55 y 0,56, con una PSNR de 0,62. Todos los modelos propuestos tuvieron unas características de funcionamiento estadísticas superiores a la PSNR en 7 de los 13 experimentos. En base a cada métrica, cada modelo de VGA con RR se situó en el grupo de los modelos de mejores prestaciones el número de veces siguientes:

Estadística	Yonsei RR10k	Yonsei RR64k	Yonsei RR128k	PSNR
Correlación	13	13	13	7
RMSE	13	13	13	6
Relación de afloramiento	13	13	13	10

Las correlaciones medias del análisis básico de los modelos de CIF con RR tuvieron todas el valor 0,78, con una PSNR de 0,66. Las correlaciones del modelo individual de algunos experimentos fueron de hasta 0,90. El RMSE medio de los modelos de CIF con RR fueron todos de 0,59, con una PSNR de 0,72. La relación de afloramiento media de los modelos de CIF con RR fue de 0,51 y 0,52, y la PSNR de 0,63. Todos los modelos propuestos tuvieron unas características de funcionamiento estadísticas superiores a la PSNR en 10 de los 14 experimentos. En base a cada métrica, cada modelo de CIF con RR se situó en el grupo de los modelos de mejores prestaciones el número de veces siguientes:

Estadística	Yonsei RR10k	Yonsei RR64k	PSNR
Correlación	14	14	5
RMSE	14	14	4
Relación de afloramiento	14	14	5

Las correlaciones medias del análisis básico de los modelos de QCIF con RR tomaron los valores de 0,77 y 0,79, con una PSNR de 0,66. Las correlaciones del modelo individual de algunos experimentos fueron de hasta 0,89. El RMSE medio de los modelos de QCIF con RR fueron de 0,58 y 0,60, con una PSNR de 0,72. La relación de afloramiento media de los modelos de QCIF con RR fue de 0,49 y 0,51, y la PSNR de 0,60. Todos los modelos propuestos tuvieron unas características de funcionamiento estadísticas superiores a la PSNR en al menos 9 de los 14 experimentos. En base a cada métrica, cada modelo de QCIF con RR se situó en el grupo de los modelos de mejores prestaciones el número de veces siguientes:

Estadística	Yonsei RR1k	Yonsei RR10k	PSNR
Correlación	14	14	5
RMSE	14	14	4
Relación de afloramiento	12	13	4

Anexo 2

Métodos de referencia reducida del Modelo A***

ÍNDICE

	Página
1	Introducción..... 7
2	Modelos de referencia reducida EPSNR 8
2.1	PSNR en los bordes 8
2.2	Selección de características de secuencias de vídeo de la fuente 13
2.3	Registro espacial/temporal y ajuste de ganancia/desplazamiento 16
2.4	Cálculo de EPSNR y postproceso..... 19
2.5	Anchura de banda óptima del canal lateral 20
	Apéndice 1 21

1 Introducción

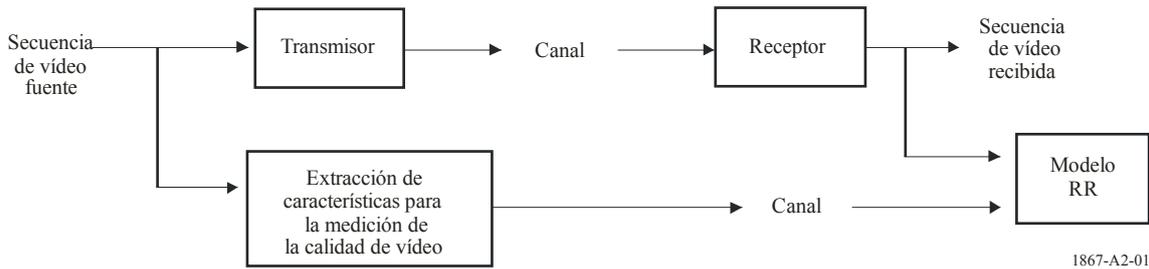
Aunque la relación señal a ruido de cresta (PSNR) se ha utilizado ampliamente como una medición objetiva de la calidad de vídeo, también se ha informado de que no representa adecuadamente la calidad de percepción de vídeo. Analizando cómo perciben los humanos la calidad de vídeo, se observa que el sistema visual humano es sensible a la degradación de los bordes. En otras palabras, cuando los píxeles de los bordes de las imágenes de un vídeo se difuminan, los evaluadores tienden a otorgar calificaciones bajas al vídeo, aunque la PSNR sea elevada. En base a estas observaciones, se han desarrollado modelos de referencia reducida que miden principalmente las degradaciones en los bordes.

La Fig. 1 ilustra el funcionamiento de un modelo de referencia reducida. Las características que se utilizan para la medición de la calidad de vídeo en un punto de supervisión se extraen de la secuencia de vídeo fuente y se transmiten. En el Cuadro 5 se muestran las anchuras de banda de los canales laterales para las características seleccionadas, que se han verificado en la prueba VQEG MM.

*** Este modelo es idéntico al especificado en el Anexo A a la Recomendación UIT-T J.246.

FIGURA 1

Diagrama de bloques del modelo de referencia reducida



CUADRO 5

Anchuras de banda de los canales laterales

Formato de vídeo	Anchuras de banda probadas
QCIF	1 kbit/s, 10 kbit/s
CIF	10 kbit/s, 64 kbit/s
VGA	10 kbit/s, 64 kbit/s, 128 kbit/s

2 Modelos de referencia reducida EPSNR

2.1 PSNR del borde

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

Puede utilizarse un algoritmo de detección de borde, aunque generalmente las diferencias en los resultados son pequeñas. Por ejemplo, para localizar los píxeles del borde puede utilizarse un operador gradiente. Se han propuesto varios operadores gradiente. En muchos algoritmos de detección de bordes, se calcula mediante operadores gradiente en primer lugar la imagen del gradiente horizontal $g_{horizontal}(m,n)$ y la imagen del gradiente vertical $g_{vertical}(m,n)$. A continuación, se calcula la imagen de gradiente de magnitud $g(m,n)$ del modo siguiente:

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

Por último, se aplica una operación de fijación de umbrales a la imagen del gradiente de magnitud $g(m,n)$ para determinar los píxeles del borde. Dicho de otro modo, los píxeles cuyos gradientes de magnitud superen un cierto valor umbral se consideran píxeles de borde.

Las Figs. 2 a 6 ilustran dicho procedimiento. La Fig. 2 muestra una imagen fuente. La Fig. 3 muestra una imagen de gradiente horizontal $g_{horizontal}(m,n)$, obtenida por aplicación de un operador gradiente horizontal a la imagen fuente de la Fig. 2. La Fig. 4 muestra una imagen de gradiente vertical $g_{vertical}(m,n)$, obtenida aplicando un operador gradiente vertical a la imagen fuente de

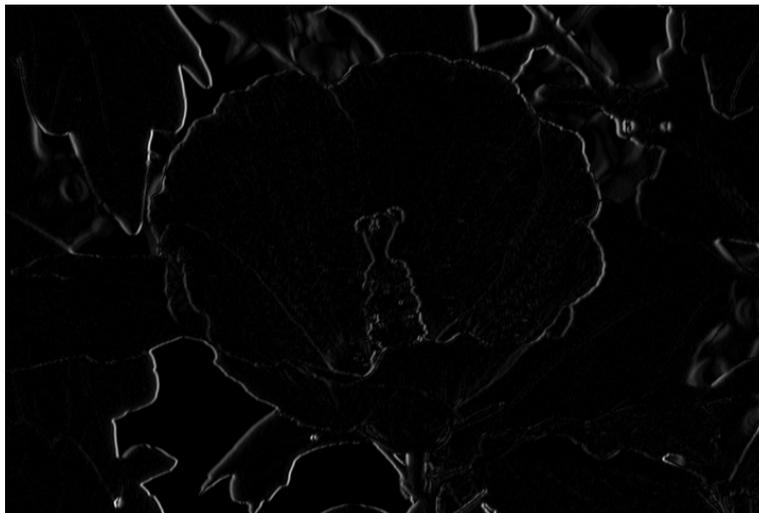
la Fig. 2. La Fig. 5 muestra la imagen del gradiente de magnitud (imagen de borde), mientras que la Fig. 6 muestra la imagen de bordes binarios (imagen máscara) obtenida por aplicación de umbrales a la imagen de gradiente de magnitud de la Fig. 5.

FIGURA 2

Imagen fuente (imagen original)

1867-A2-02

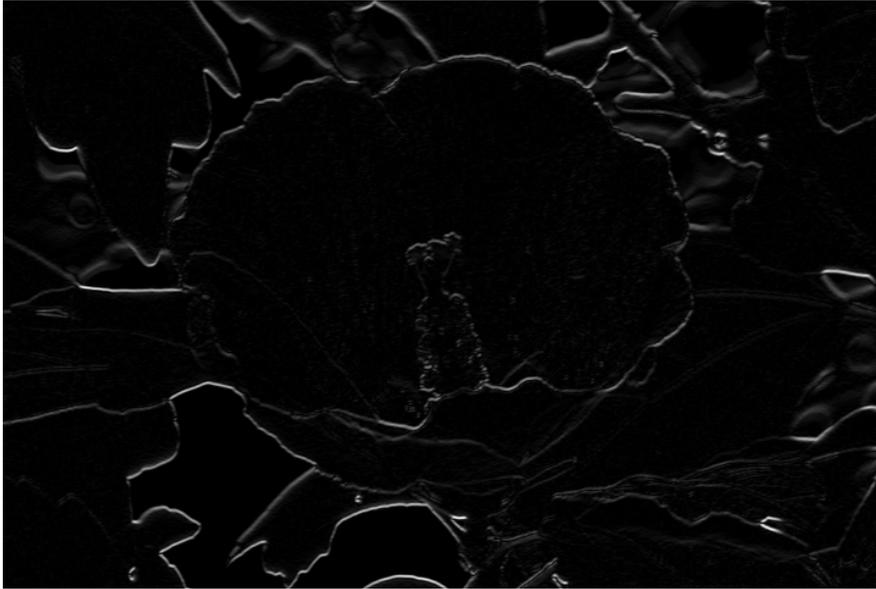
FIGURA 3

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

1867-A2-03

FIGURA 4

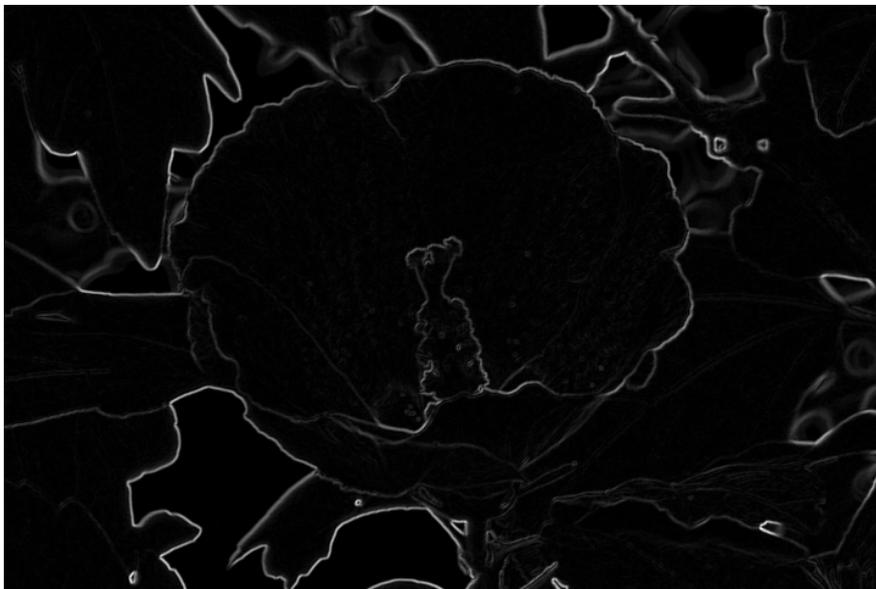
Imagen de gradiente vertical obtenida aplicando un operador gradiente vertical a la imagen fuente de la Fig. 2



1867-A2-04

FIGURA 5

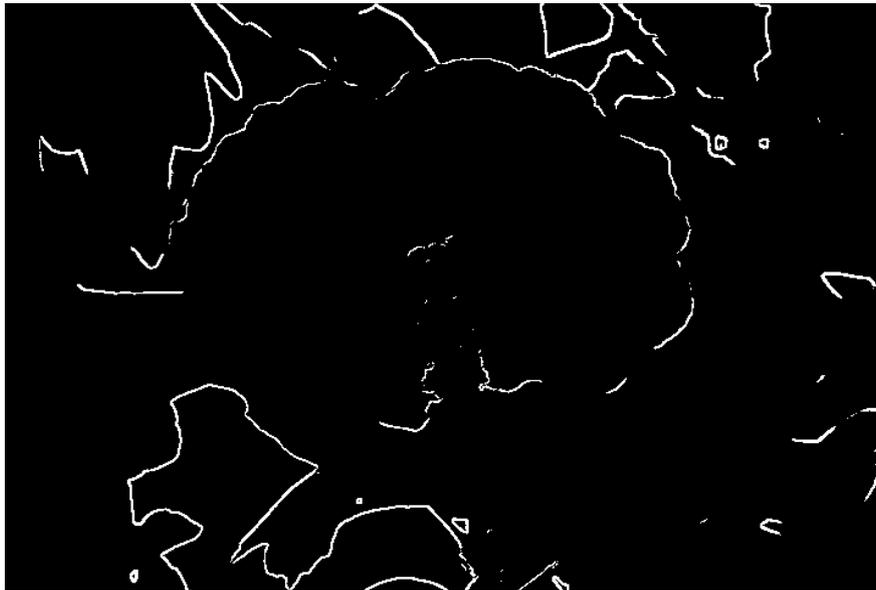
Imagen de gradiente magnitud



1867-A2-05

FIGURA 6

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

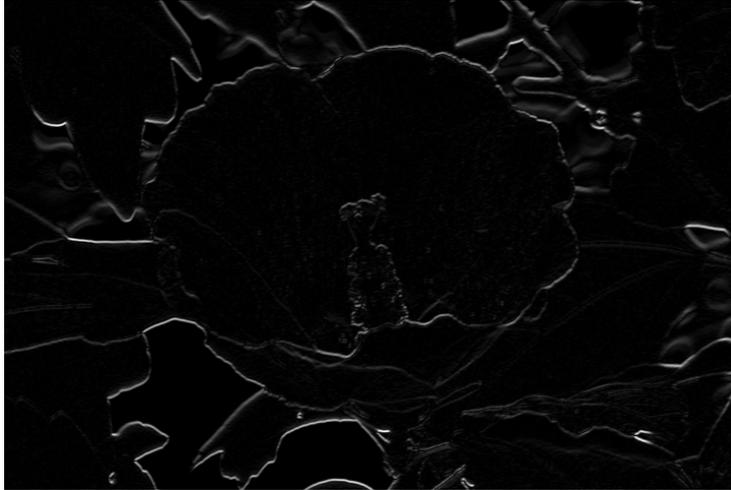


1867-A2-06

Alternativamente, puede utilizarse un procedimiento modificado para determinar los píxeles de los bordes. Por ejemplo, en primer lugar puede aplicarse un operador gradiente vertical a la imagen fuente, obteniendo así una imagen de gradiente vertical. A continuación se aplica un operador gradiente horizontal a la imagen de gradiente vertical, produciendo una imagen de gradiente sucesivo modificada (imagen de gradiente horizontal y vertical). Finalmente, pueden fijarse umbrales a la imagen de gradiente sucesivo modificada para determinar los píxeles de los bordes. Dicho de otro modo, los píxeles de la imagen de gradiente sucesivo modificada que superen el umbral se consideran píxeles de bordes. En las Figs. 7 a 9 se ilustra el procedimiento modificado. La Fig. 7 muestra una imagen de gradiente vertical $g_{vertical}(m,n)$, obtenida por aplicación de un operador 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), obtenida por aplicación de un operador gradiente horizontal a la imagen de gradiente vertical de la Fig. 7. En la Fig. 9 se muestra la imagen de bordes binarios (imagen máscara) obtenida por aplicación de una operación de determinación de umbrales a la imagen de gradiente sucesivo modificada de la Fig. 8.

FIGURA 7

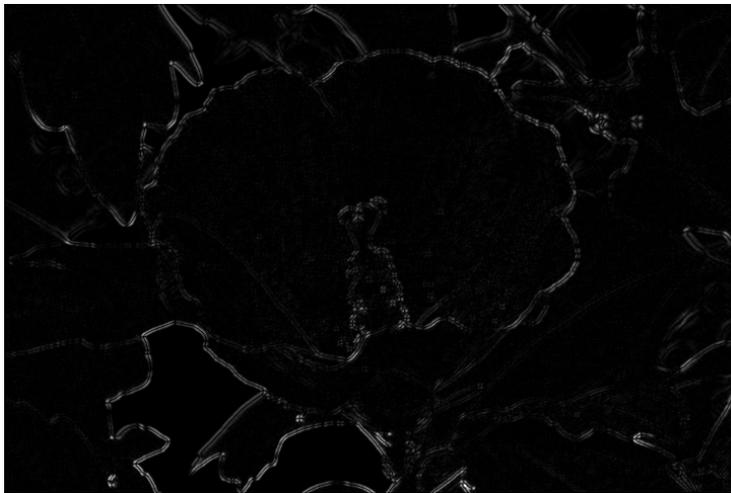
Imagen de gradiente vertical obtenida aplicando un operador gradiente vertical a la imagen fuente de la Fig. 2



1867-A2-07

FIGURA 8

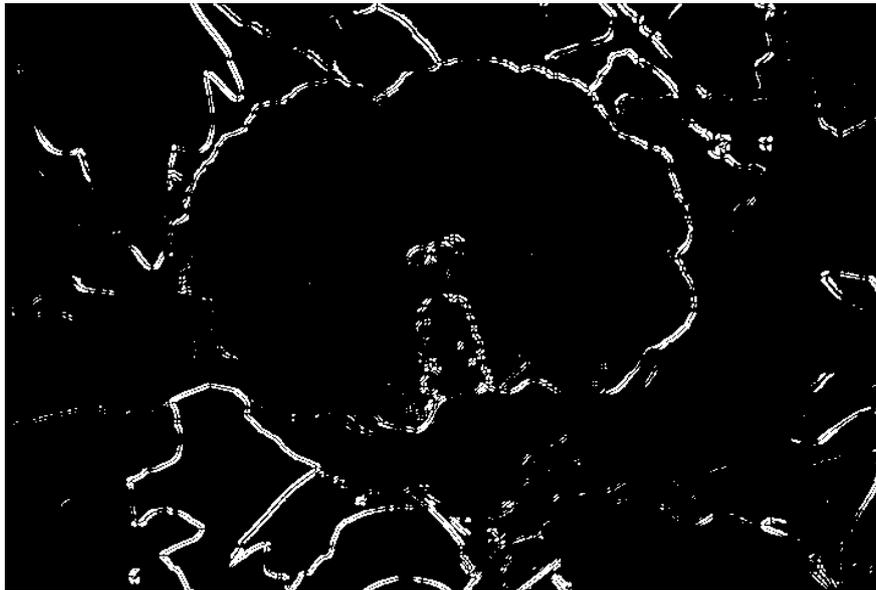
Imagen de gradiente sucesiva modificada (imagen de gradiente horizontal y vertical), obtenida aplicando un operador gradiente horizontal a la imagen de gradiente vertical de la Fig. 7



1867-A2-08

FIGURA 9

Imagen de borde binaria (imagen máscara) obtenida aplicando un umbral a la imagen de gradiente sucesiva modificada de la Fig. 8



1867-A2-09

Nótese que ambos métodos pueden considerarse algoritmos de detección de bordes. Se puede escoger cualquier algoritmo de detección de bordes en función de la naturaleza de los vídeos y de los algoritmos de compresión. No obstante, unos métodos pueden funcionar mejor que otros.

Por tanto, en el modelo se aplica en primer lugar un operador de detección de bordes, produciendo las imágenes de bordes (Figs. 5 y 8). A continuación, se genera una imagen de máscara (imagen de bordes binarios) por aplicación de umbrales a la imagen de bordes (Figs. 6 y 9). Dicho de otro modo, los píxeles de la imagen de bordes cuyo valor sea inferior al umbral t_e se ponen a cero y aquellos cuyo valor sea igual o mayor que el umbral se ponen a un valor no nulo. Las Figs. 6 y 9 muestran ejemplos de imágenes máscara. Dado que un vídeo puede contemplarse como una secuencia de cuadros o tramas, el procedimiento descrito puede aplicarse a cada uno de los cuadros o tramas de los vídeos. Como el modelo puede utilizarse con vídeos basados en tramas o con vídeos basados en cuadros, el término «imagen» se utilizará para indicar una trama o un cuadro.

2.2 Selección de características de secuencias de vídeo fuente

Dado que se trata de un modelo de referencia reducida (RR), es necesario extraer un conjunto de características de cada imagen de la secuencia de vídeo fuente. En el modelo EPSNR de RR, se selecciona un cierto número de píxeles de bordes de cada imagen. A continuación se codifican y transmiten la ubicación y el valor de cada uno de los píxeles. Sin embargo, en algunas secuencias de vídeo el número de píxeles de bordes puede ser muy pequeño tras la aplicación de un umbral. En el caso peor, dicho número puede ser cero (imágenes borradas o imágenes de muy baja frecuencia). Para abordar este problema, si el número de píxeles de bordes es menor que un valor determinado, el usuario puede reducir el umbral hasta que el número de píxeles de bordes sea superior a un valor dado. Alternativamente, pueden seleccionarse píxeles de bordes que correspondan a los valores mayores de la imagen de gradiente horizontal y vertical. Cuando en un cuadro no haya píxeles de bordes (imágenes borradas), puede seleccionarse aleatoriamente el número de píxeles o eliminar el cuadro. Por ejemplo, si deben seleccionarse 10 píxeles de bordes de cada cuadro, los píxeles de la imagen de gradiente horizontal y vertical pueden ordenarse según sus valores, y seleccionar los 10 valores mayores. No obstante, este procedimiento puede generar varios píxeles de borde en los

mismos puntos. Para resolver este problema, puede seleccionarse el número deseado de píxeles de la imagen de gradiente horizontal y vertical y seleccionar aleatoriamente el número deseado de píxeles de bordes entre los píxeles seleccionados de dicha imagen. En los modelos probados en las pruebas de multimedios del VQEG, el número deseado de píxeles de bordes se obtuvieron aleatoriamente entre un conjunto numeroso de píxeles de bordes. El conjunto de píxeles de bordes se obtiene por aplicación de una operación umbral a la imagen gradiente.

En los modelos EPSNR de RR, se codifica la situación y valor de los píxeles de bordes. Nótese que durante la codificación, puede aplicarse un recorte selectivo (*cropping*) de la imagen. Para evitar que se seleccionen píxeles de bordes en las zonas recortadas, el modelo selecciona píxeles de bordes en la zona central de la imagen (Fig. 10). El Cuadro 6 muestra los tamaños de la imagen después del recorte, así como el número de bits necesarios para la codificación de la situación y valor de píxel de cada píxel de borde.

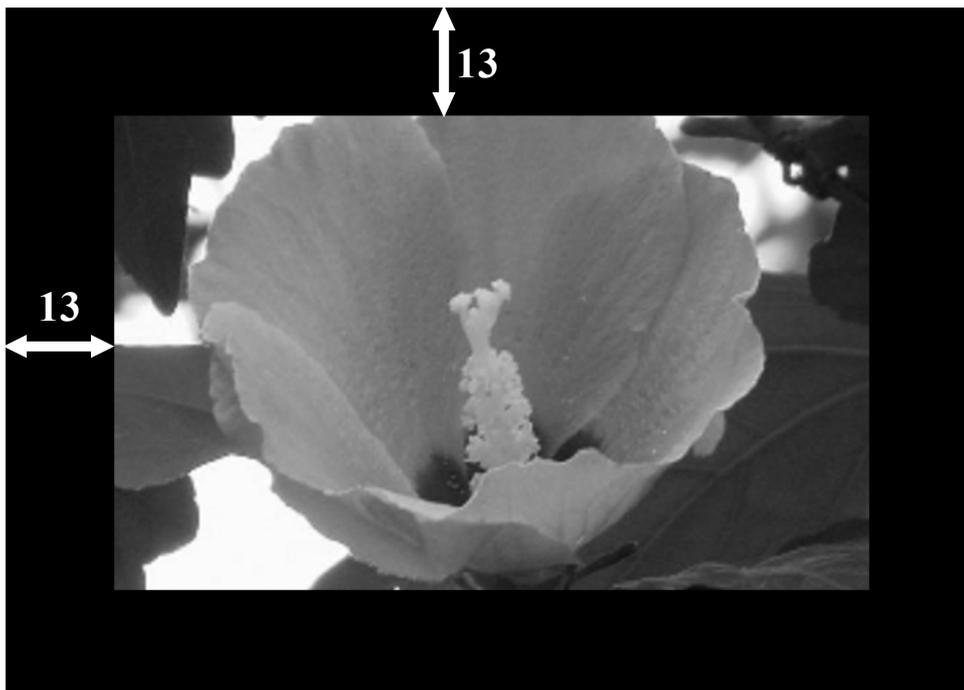
CUADRO 6

Bits requeridos para cada píxel de borde

Formato de vídeo	Tamaño	Tamaño después del recorte	Bits para determinar la situación	Bits para cada valor de píxel	Número total de bits por píxel
QCIF	176 × 144	168 × 136	15	8	23
CIF	352 × 288	338 × 274	17	8	25
VGA	640 × 480	614 × 454	19	8	27

FIGURA 10

Ejemplo de recorte (VGA) y la zona central seleccionada



El modelo selecciona píxeles de bordes de cada cuadro según la anchura de banda permitida (Cuadro 5). Las Figs. 7 y 8 muestran el número de píxeles de bordes por cuadro que pueden transmitirse en las anchuras de banda probadas.

CUADRO 7

Número de píxeles de bordes por cuadro (30 cuadros/s)

Formato de vídeo	1 kbit/s	10 kbit/s	64 kbit/s	128 kbit/s
QCIF	1	14		
CIF		13	85	
VGA		12	79	158

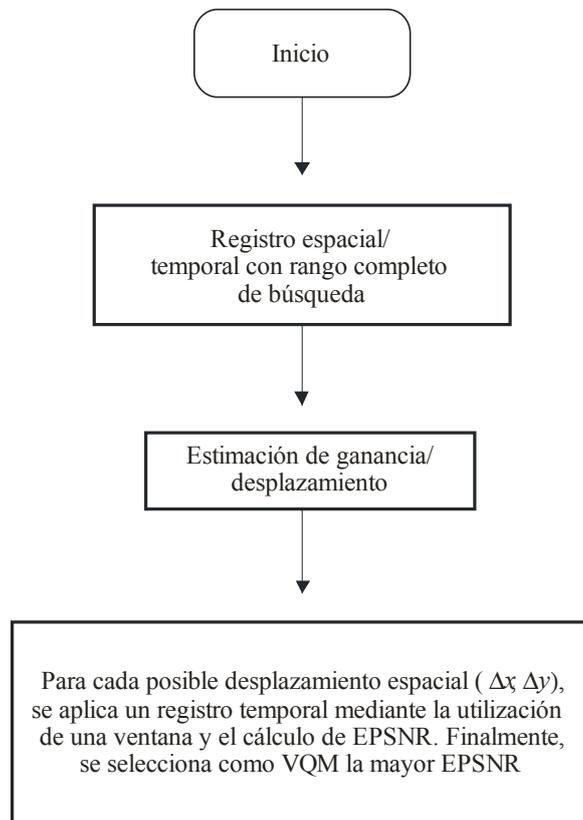
CUADRO 8

Número de píxeles de bordes por cuadro (25 cuadros /s)

Formato de vídeo	1 kbit/s	10 kbit/s	64 kbit/s	128 kbit/s
QCIF	1	17		
CIF		16	102	
VGA		14	94	189

FIGURA 11

Diagrama de flujos del modelo



2.3 Registro espacial y temporal y ajuste de ganancia y desplazamiento

Antes del cálculo de la diferencia entre los píxeles de bordes de la secuencia de vídeo fuente y los de la secuencia de vídeo procesada o secuencia de vídeo recibida en el receptor, el modelo aplica un registro espacial/temporal y un ajuste a la ganancia y el desplazamiento. En primer lugar se aplica un algoritmo de búsqueda completo, para determinar los desplazamientos espaciales y temporales globales así como los valores de ajuste de ganancia y desplazamiento (Fig. 11). Para cada posible desplazamiento espacial $(\Delta x, \Delta y)$, se hace un registro temporal y se calcula la EPSNR. Finalmente, se selecciona como métrica de la calidad de vídeo (VQM, *video quality metric*) el mayor valor de EPSNR.

En el punto de supervisión, la secuencia de vídeo procesada se alinea con los píxeles de bordes extraídos de la secuencia de vídeo fuente. Si la anchura de banda del canal lateral es pequeña, sólo habrá disponibles unos pocos píxeles de bordes de la secuencia de vídeo fuente (Fig. 12). En consecuencia, el registro temporal puede ser inexacto si se calcula en base a un solo cuadro (Fig. 13). Para abordar este problema, el modelo utiliza una ventana para el registro temporal. En lugar de utilizar un solo cuadro de la secuencia de vídeo procesada, el modelo genera una ventana que consta de un número de cuadros adyacentes para determinar el desplazamiento temporal óptimo. La Fig. 14 ilustra el procedimiento. El error cuadrático medio en la ventana se calcula de la forma siguiente:

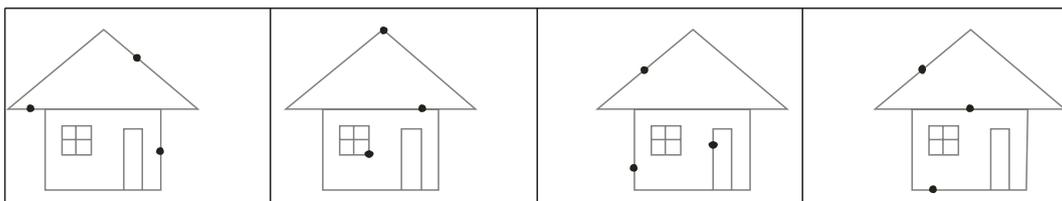
$$MSE_{window} = \frac{1}{N_{win}} \sum (E_{SRC}(i) - E_{PVS}(i))^2$$

donde MSE_{window} es el error cuadrático medio de la ventana, $E_{SRC}(i)$ es un píxel de borde en la ventana que tiene su correspondiente píxel en la secuencia de vídeo procesada, $E_{PVS}(i)$ es el píxel de la secuencia de vídeo procesada correspondiente al píxel de borde, y N_{win} es el número total de píxeles de bordes utilizados para calcular MSE_{window} . El error cuadrático medio de la ventana se utiliza como la diferencia entre un cuadro de la secuencia de vídeo procesada y el correspondiente cuadro de la secuencia de vídeo fuente.

El tamaño de la ventana puede determinarse teniendo en cuenta la naturaleza de la secuencia de vídeo procesada. Para una aplicación típica, se recomienda una ventana de dos segundos. Alternativamente, pueden aplicarse diversos tamaños de ventana y utilizar la que presente el error cuadrático medio más pequeño.

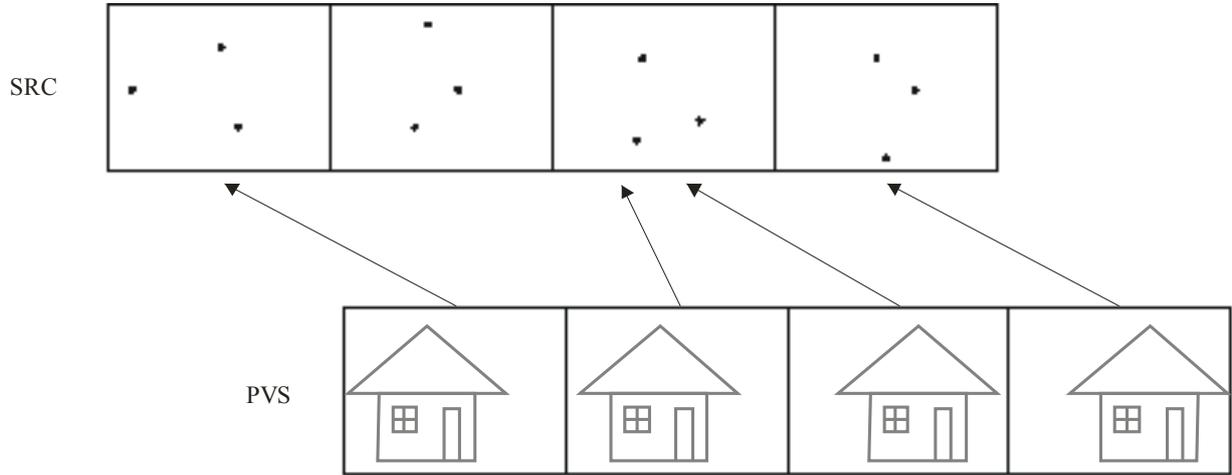
FIGURA 12

Selección de píxeles de bordes de la secuencia de vídeo fuente



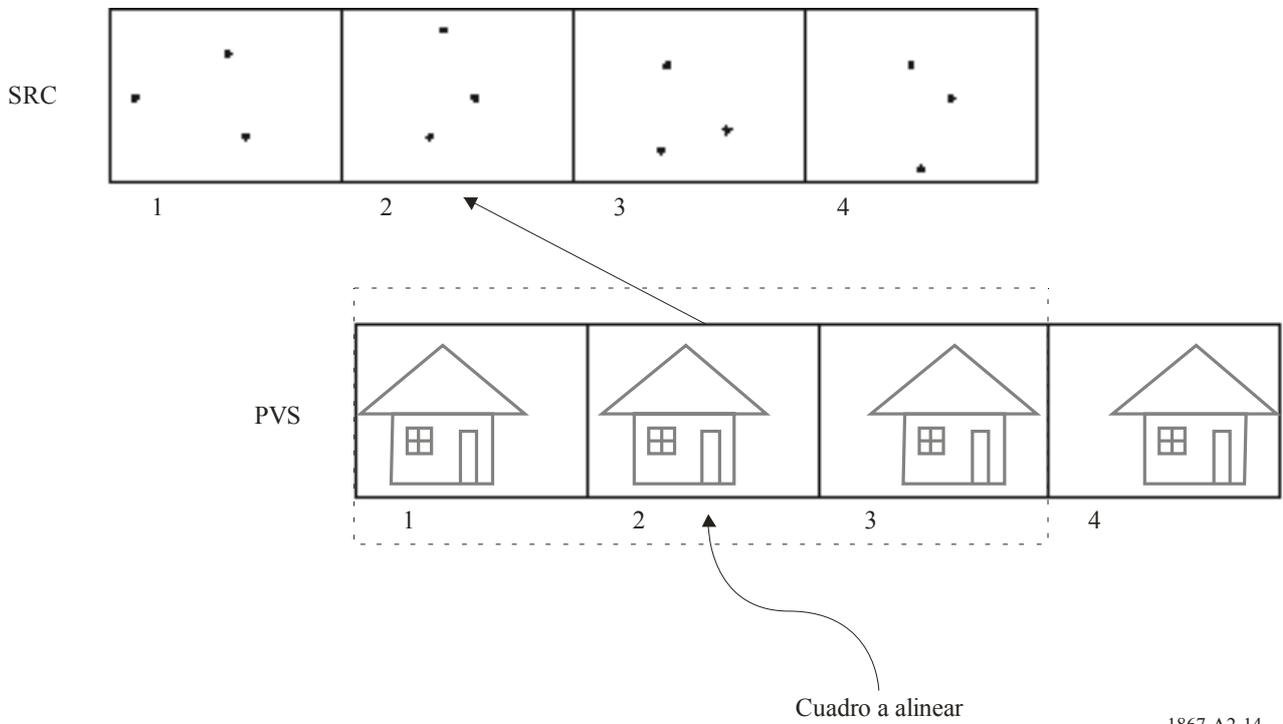
1867-A2-12

FIGURA 13
Alineación de la secuencia de vídeo procesada con los píxeles de bordes de la secuencia de vídeo fuente



1867-A2-13

FIGURA 14
Alineación de la secuencia de vídeo procesada con los píxeles de bordes al utilizar una ventana



1867-A2-14

Cuando la secuencia de vídeo se codifica con relaciones de compresión elevadas, el codificador puede reducir el número de cuadros por segundo y en ese caso la secuencia de vídeo procesada incluye cuadros repetidos (Fig. 15). En la Fig. 15, la secuencia de vídeo procesada no tiene algunos de los cuadros de la secuencia de vídeo fuente (Cuadros 2, 4, 6 y 8). En este caso, el modelo no utiliza cuadros repetidos para calcular el error cuadrático medio. Dicho de otra forma, el modelo realiza el registro temporal utilizando el primer cuadro (cuadro válido) de cada bloque repetido. Por tanto, en la Fig. 16, para el registro temporal sólo se utilizan tres cuadros de la ventana (3, 5 y 7).

FIGURA 15

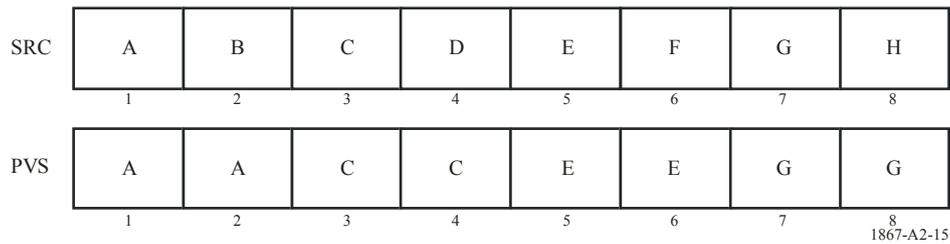
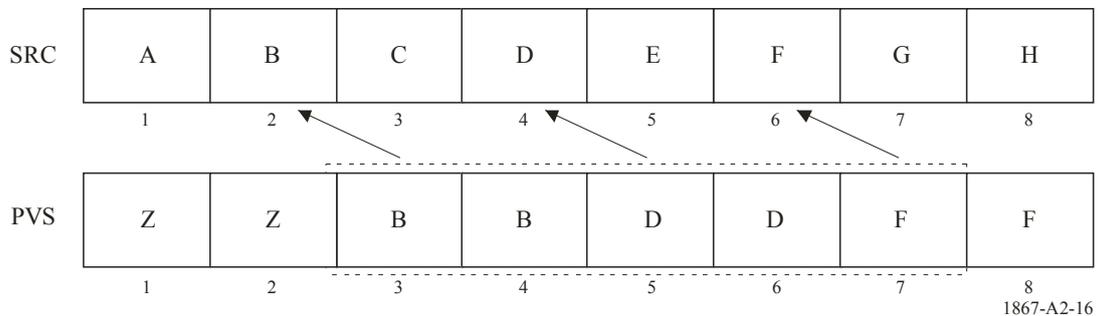
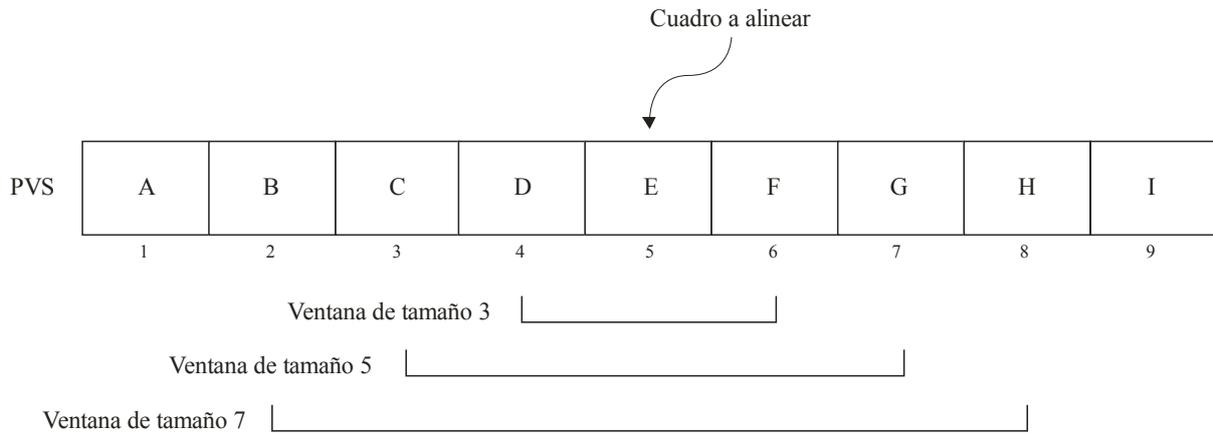
Ejemplo de cuadros repetidos

FIGURA 16

Tratamiento de los cuadros repetidos

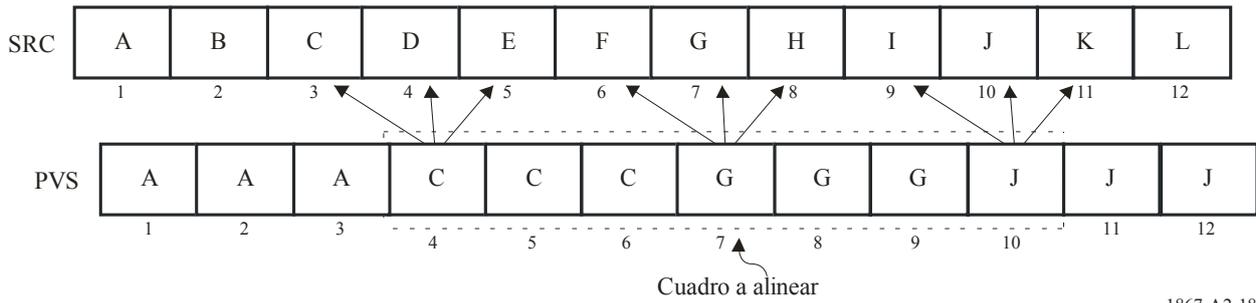
Es posible que una secuencia de vídeo procesada tenga una repetición de tramas irregular, lo cual puede hacer que el método de registro temporal que utiliza una ventana produzca resultados inexactos. Para abordar este problema, es posible ajustar localmente cada cuadro de la ventana dentro de un rango de valores dado (por ejemplo, ± 1), tal como se muestra en la Fig. 18 después de un registro temporal que utiliza una ventana. En ese caso, para calcular la EPSNR se utiliza el ajuste local que proporciona el error cuadrático medio (MSE) mínimo.

FIGURA 17
Ventanas de varios tamaños



BT.1867-17

FIGURA 18
Ajuste local para el registro temporal que utiliza una ventana



1867-A2-18

2.4 Cálculo de EPSNR y postprocesamiento

Una vez realizado el registro temporal, se calcula la media de las diferencias entre los píxeles de bordes de la secuencia de vídeo fuente y los correspondientes píxeles de la secuencia de vídeo procesada, que puede interpretarse como el error cuadrático medio de los bordes de la secuencia de vídeo procesada (MSE_{edge}). Finalmente, la EPSNR (PSNR de los bordes) se calcula de la forma siguiente:

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

donde:

p es el valor de cresta de la imagen.

En la codificación de vídeo multimedios, los cuadros pueden repetirse debido a velocidades de cuadro reducidas y pueden congelarse debido a errores de transmisión, todo lo cual degrada la calidad percibida del vídeo. Para dar un tratamiento adecuado a este efecto, el modelo aplica el ajuste siguiente antes del cálculo de la EPSNR:

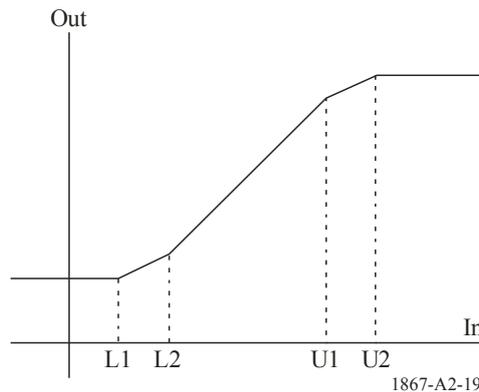
$$MSE_{frozen_frame_considered} = MSE_{edge} \times \frac{K \times N_{total_frame}}{N_{total_frame} - N_{total_frozen_frame}}$$

donde $MSE_{frozen_frame_considered}$ es el error cuadrático medio que tiene en cuenta los cuadros repetidos y congelados, N_{total_frame} es el número total de cuadros, $N_{total_frozen_frame}$ es el número total de cuadros congelados y K es una constante. En el modelo verificado en la prueba multimedios del VQEG, K se fijó en 1.

Cuando la EPSNR supera un determinado valor, la calidad percibida se satura. En este caso, es posible fijar un límite superior de la EPSNR. Además, cuando es deseable que exista una relación lineal entre la EPSNR y la DMOS (nota media de opinión de diferencias), puede aplicarse una función lineal por tramos, tal como se ilustra en la Fig. 19. En el modelo verificado en la prueba multimedios del VQEG, sólo se fijó el límite superior a 50 pues se aplicó un ajuste mediante curva polinomial.

FIGURA 19

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



2.5 Anchura de banda óptima para canal lateral

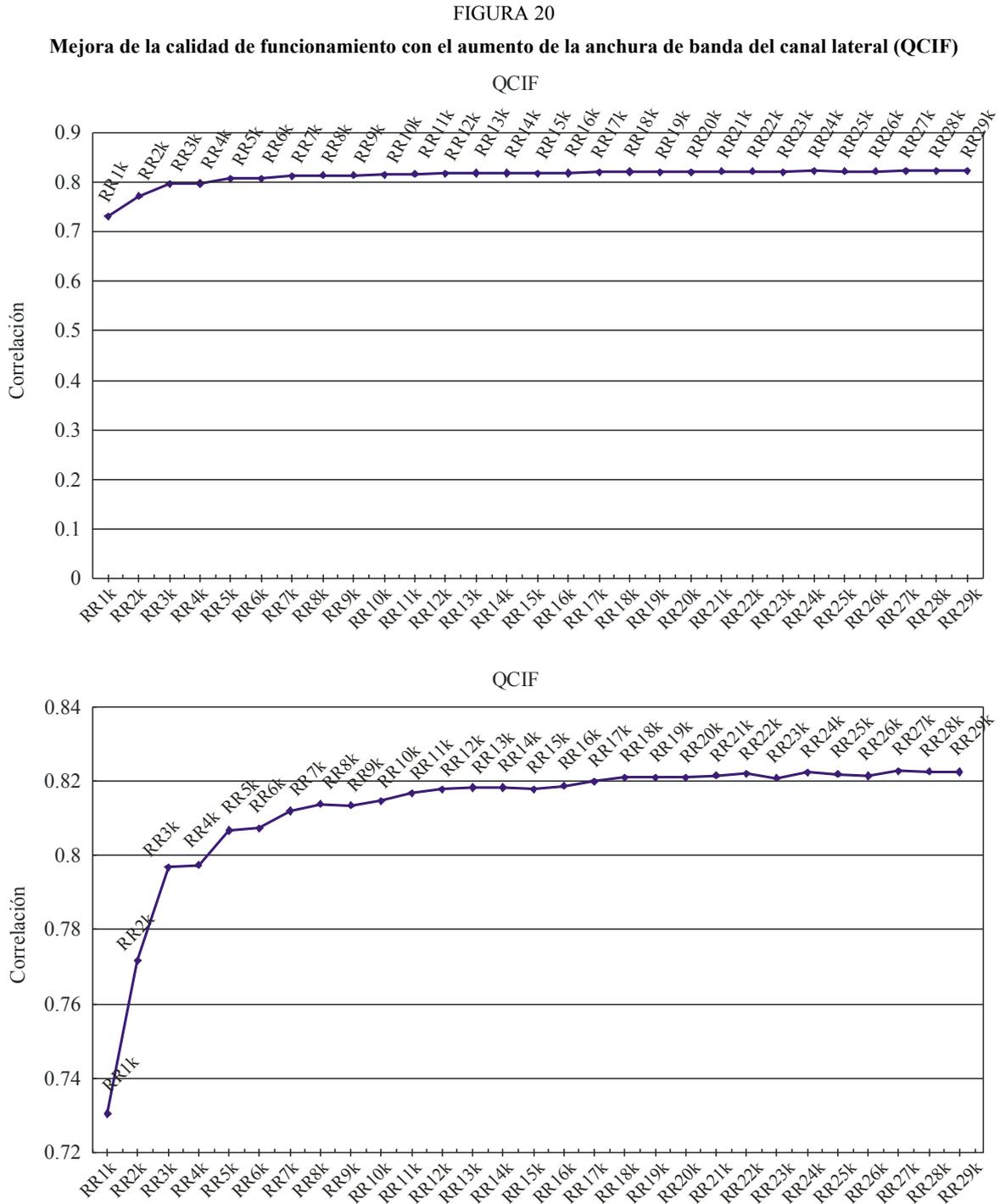
En el Apéndice se muestra la comparación de la calidad de funcionamiento cuando se aumenta la anchura de banda del canal lateral. Para el formato QCIF, se observa que los coeficientes de correlación están casi saturados a aproximadamente unos 10 kbit/s. A partir de ese nivel, un aumento de la anchura de banda produce una mejora de aproximadamente el 1%. Para el formato CIF, se observa que los coeficientes de correlación están prácticamente saturados a unos 15 kbit/s. A partir de ese nivel, un aumento de la anchura de banda produce una mejora de aproximadamente el 0,5%. Para el formato VGA, se observa que los coeficientes de correlación están prácticamente saturados a unos 30 kbit/s. A partir de ese nivel, un aumento de la anchura de banda produce una mejora de aproximadamente el 0,5%.

Los modelos de EPSNR de referencia reducida para la medida objetiva de la calidad de vídeo se basan en la medición de la degradación en los bordes. Los modelos pueden implementarse en tiempo real con un uso moderado de potencia de cálculo. Los modelos están bien adaptados a aplicaciones que requieran la supervisión de la calidad del vídeo en tiempo real y en las que existan canales laterales.

Apéndice 1 al Anexo 2

1 Anchuras de banda del canal lateral óptimas

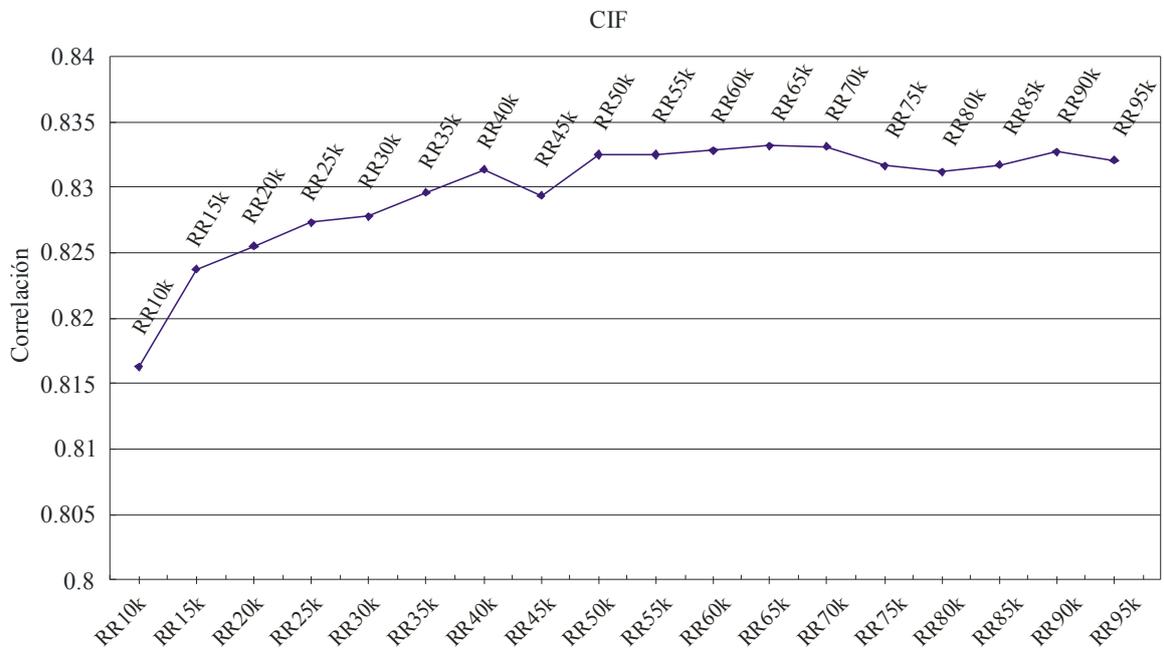
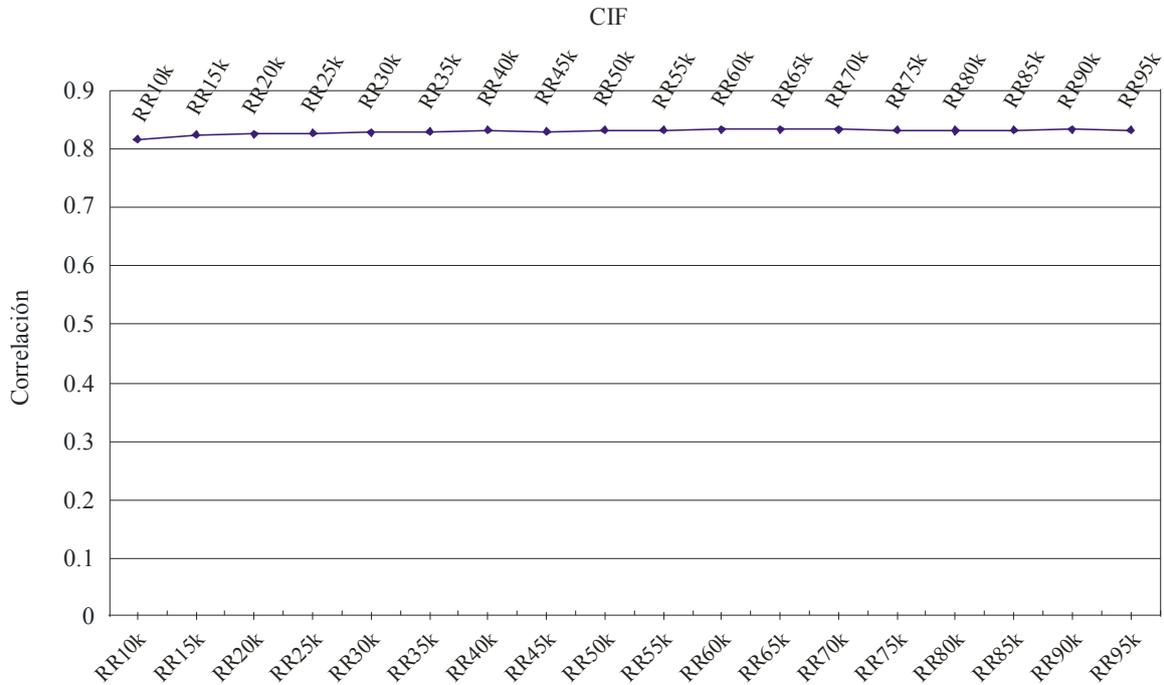
La Fig. 20 muestra coeficientes de correlación para distintas anchuras de banda del canal lateral para los conjuntos de vídeos QCIF. Puede observarse que los coeficientes de correlación están casi saturados a unos 10 kbit/s. A partir de ese nivel, un aumento de la anchura de banda produce una mejora de aproximadamente el 1%.



La Fig. 21 muestra coeficientes de correlación para distintas anchuras de banda del canal lateral para los conjuntos de vídeos CIF. Puede observarse que los coeficientes de correlación están casi saturados a unos 15 kbit/s. A partir de ese nivel, un aumento de la anchura de banda produce una mejora de aproximadamente el 0,5%.

FIGURA 21

Mejora de la calidad de funcionamiento con el aumento de la anchura de banda del canal lateral (CIF)



La Fig. 22 muestra coeficientes de correlación para distintas anchuras de banda del canal lateral para los conjuntos de videos VGA. Puede observarse que los coeficientes de correlación están casi saturados a unos 30 kbit/s. A partir de ese nivel, un aumento de la anchura de banda produce una mejora de aproximadamente el 0,5%.

FIGURA 22

Mejora de la calidad de funcionamiento con el aumento de la anchura de banda del canal lateral (VGA)

