

## RECOMENDACIÓN UIT-R BT.1683

**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**

(Cuestión UIT-R 44/6)

(2004)

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 los métodos objetivos convencionales ya no resultan plenamente adecuados para la medición de la calidad de vídeo percibida de los sistemas de vídeo digital con compresión;
- c) que las mediciones objetivas de la calidad de vídeo percibida complementarán los métodos de prueba objetiva convencionales;
- d) que los actuales métodos formales de evaluación subjetiva son lentos y costosos y no suelen estar adaptados a las condiciones de funcionamiento;
- e) que la medición objetiva de la calidad de vídeo percibida puede resultar útil como complemento de los métodos de evaluación subjetiva,

*recomienda*

- 1** se utilicen las directrices, objeto y restricciones que figuran en el Anexo 1 al aplicar los modelos de calidad de vídeo objetiva que figuran en los Anexos 2-5;
- 2** se utilicen los modelos de calidad de vídeo objetiva de los Anexos 2-5 para la medición objetiva de la calidad de vídeo percibida.

**Anexo 1****Resumen**

En la presente Recomendación se especifican métodos para estimar la calidad de vídeo percibida de un sistema de transmisión de vídeo unidireccional. Esta Recomendación es aplicable a las señales en banda base. Los métodos de estimación de esta Recomendación son aplicables a:

- la evaluación, especificación y pruebas de aceptación de códecs;
- la supervisión potencial de la calidad en servicio, en tiempo real, en la fuente;
- la supervisión remota de la calidad del destino cuando hay disponible una copia de la fuente;
- la medición de la calidad de un sistema de almacenamiento de vídeo o de un sistema digital que utilice técnicas de compresión y descompresión de vídeo, ya sea en un solo paso o mediante una combinación de estas técnicas.

## Introducción

La capacidad de medir automáticamente la calidad del vídeo de radiodifusión se ha considerado desde hace tiempo de enorme interés para la industria. La industria de la radiodifusión necesita que estos instrumentos sustituyan o complementen las costosas y lentas pruebas de la calidad subjetiva. Tradicionalmente, las mediciones de la calidad objetiva se han obtenido mediante cálculo de la relación señal/ruido de cresta (PSNR). A pesar de su utilidad como indicador de la calidad, se ha demostrado que la PSNR es una representación poco satisfactoria de la calidad perceptual. Para superar las limitaciones asociadas a la PSNR, se han llevado a cabo investigaciones para definir algoritmos que permitan medir la calidad perceptual del vídeo de radiodifusión. Estos instrumentos de medición de la calidad perceptual objetiva pueden aplicarse a las pruebas de calidad de funcionamiento de una red de radiodifusión, como ayuda en la adquisición de equipos y para el desarrollo de nuevas técnicas de codificación del vídeo de radiodifusión. En estos últimos años se han dedicado ingentes esfuerzos al desarrollo de instrumentos fiables y precisos que puedan utilizarse para medir con objetividad la calidad perceptual del vídeo de radiodifusión. En la presente Recomendación se definen modelos computacionales objetivos que han demostrado ser mejores que la PSNR como instrumentos de medida automáticos para la evaluación de la calidad del vídeo de radiodifusión. Estos modelos se probaron con material de 525 líneas y de 625 líneas conforme con la Recomendación UIT-R BT.601, característico de la distribución secundaria de vídeo de calidad de televisión con codificación digital.

La calidad de funcionamiento de los modelos de calidad perceptual se determinó mediante dos evaluaciones paralelas de material de vídeo de prueba<sup>1</sup>. En la primera evaluación se utilizó el método de doble estímulo con escala de calidad continua (DSCQS) subjetivo estándar para obtener los índices subjetivos de la calidad del material de vídeo determinados por equipos de observadores (Recomendación UIT-R BT.500 – Metodología para la evaluación subjetiva de la calidad de las imágenes de televisión). En la segunda evaluación, se obtuvieron los índices objetivos a partir de modelos computacionales objetivos. Para cada modelo se computaron varias métricas a fin de medir la precisión y coherencia con la que los índices objetivos predecían los subjetivos. La parte de la prueba correspondiente a la evaluación subjetiva se llevó a cabo en tres laboratorios independientes. Dos de éstos, Communications Research Center (CRC, Canadá) y Verizon (Estados Unidos de América), realizaron la prueba utilizando secuencias de 525/60 Hz, mientras que el tercero, Fondazione Ugo Bordoni (FUB, Italia), realizó la prueba utilizando secuencias de 625/50 Hz. Varios laboratorios «proponentes» elaboraron modelos computacionales objetivos de la calidad de vídeo a partir de las mismas secuencias de vídeo ensayadas con observadores humanos por CRC, Verizon y FUB. Los resultados de las pruebas se presentan en el Apéndice 1.

La presente Recomendación contiene los modelos computacionales objetivos relacionados en el Cuadro 1.

En los Anexos 2-5 se facilita una descripción completa de los cuatro modelos computacionales objetivos citados.

Pueden seguir utilizándose los equipos de prueba de la calidad de vídeo existentes mientras no se disponga de equipos de prueba que implementen alguno de los cuatro modelos anteriores.

---

<sup>1</sup> Doc. UIT-R 6Q/14 [septiembre de 2003] Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II (FR-TV2).

CUADRO 1

Número del modelo	Nombre	Proponente del Grupo de Expertos en Calidad de Vídeo (VQEG)	País	Anexo
1	British Telecom	D	Reino Unido	2
2	Yonsei University/Radio Research Laboratory/SK Telecom	E	Corea (Rep. de)	3
3	Centro de Investigación y Desarrollo en Telecomunicaciones (CPqD)	F	Brasil	4
4	National Telecommunications and Information Administration (NTIA)/Institute for Telecommunication Sciences (ITS)	H	Estados Unidos de América	5

Para que un modelo pueda considerarse incluido en la sección normativa de la presente Recomendación en el futuro, debe ser verificado por un organismo abierto e independiente (tal como el VQEG) que realice su evaluación técnica con sujeción a las directrices y criterios de calidad de funcionamiento establecidos por la Comisión de Estudio 6 de Radiocomunicaciones. La intención de la Comisión de Estudio 6 de Radiocomunicaciones es recomendar, finalmente, un único método de referencia normativa completa.

## 1 Objeto

En esta Recomendación se especifican métodos para estimar la calidad de vídeo percibida de un sistema de vídeo unidireccional. Esta Recomendación es aplicable a las señales en banda base. Los estimadores de la calidad de funcionamiento objetiva de vídeo se definen para la calidad extremo a extremo entre dos puntos. Los métodos de estimación se basan en el procesamiento del vídeo de componente digital de 8 bits definido en la Recomendación UIT-R BT.601<sup>2</sup>. El codificador puede utilizar varios métodos de compresión (por ejemplo el Grupo de Expertos en imágenes en movimiento (MPEG) Recomendación UIT-T H.263, etc.). Los modelos propuestos en esta Recomendación pueden utilizarse para evaluar un códec (combinación de codificador/decodificador) o una combinación de varios métodos de compresión y dispositivos de almacenamiento en memoria. Aunque en el cálculo de los estimadores de la calidad objetiva descrito en esta Recomendación se considera la degradación provocada por errores (por ejemplo, errores en los bits, paquetes rechazados), no se dispone aún de resultados de pruebas independientes para validar la utilización de estimadores en los sistemas con degradación por errores. El material de pruebas de validación no contenía errores de canal.

<sup>2</sup> Esto no impide la implementación de métodos de medida para sistemas de vídeo unidireccional que utilicen entradas y salidas de vídeo compuesto. La especificación de la conversión entre el dominio de vídeo compuesto y el de componentes no figura en la presente Recomendación. En la Norma SMPTE 170M, por ejemplo, se especifica un método para realizar esta conversión en NTSC.

## 1.1 Aplicación

En esta Recomendación se facilitan estimaciones de la calidad de vídeo para las clases de vídeo de televisión (TV0-TV3) y la clase de vídeo multimedios (MM4) definidas en el Anexo B de la Recomendación UIT-T P.911. Entre las aplicaciones de los modelos de estimación descritos en la presente Recomendación se encuentran las siguientes:

- la evaluación, especificación y pruebas de aceptación de códecs;
- la supervisión potencial de la calidad en servicio, en tiempo real, en la fuente;
- la supervisión remota de la calidad del destino cuando hay disponible una copia del origen;
- la medición de la calidad de un sistema de almacenamiento de vídeo o de un sistema digital que utilice técnicas de compresión y descompresión de vídeo, ya sea en un solo paso o mediante una combinación de estas técnicas.

## 1.2 Restricciones

Los modelos de estimación descritos en la presente Recomendación no pueden utilizarse para sustituir las pruebas subjetivas. Los valores de correlación entre dos pruebas subjetivas cuidadosamente diseñadas y ejecutadas (es decir, en dos laboratorios diferentes) suelen estar en el intervalo 0,92 a 0,97. En esta Recomendación no se facilita ningún medio de cuantificar los errores potenciales de estimación. Es conveniente que los lectores de la presente Recomendación analicen la comparación de los resultados subjetivos y objetivos disponibles para hacerse una idea del intervalo de error de estimación de los índices de calidad de vídeo.

La calidad de funcionamiento predicha por los modelos de estimación no está validada aún para los sistemas de vídeo con degradación por errores en el canal de transmisión.

## Anexo 2

### Modelo 1

#### ÍNDICE

	<i>Página</i>
1	Introducción ..... 5
2	Modelo de BTFR ..... 5
3	Detectores ..... 6
3.1	Conversión de la entrada ..... 6
3.2	Recorte y desplazamiento ..... 7
3.3	Emparejamiento ..... 8
3.3.1	Estadísticas de emparejamiento ..... 10
3.3.2	MPSNR ..... 10
3.3.3	Vectores de emparejamiento ..... 10

	<i>Página</i>
3.4	Análisis espacial de frecuencia ..... 11
3.4.1	La transformada piramidal ..... 11
3.4.2	SNR piramidal ..... 13
3.5	Análisis de textura ..... 13
3.6	Análisis de los bordes ..... 14
3.6.1	Detección de los bordes ..... 14
3.6.2	Diferenciación de los bordes ..... 14
3.7	Análisis de la MPSNR ..... 15
4	Integración ..... 15
5	Registro ..... 16
6	Referencias Bibliográficas ..... 16
Anexo 2a	..... 17

## 1 Introducción

El instrumento de evaluación automática de la calidad de vídeo de referencia completa de la BT (BTFR, *BT full-reference*) elabora predicciones de la calidad de vídeo representativas de las estimaciones humanas de la calidad. Este instrumento de medición objetiva simula digitalmente las características del sistema visual humano (HVS, *human visual system*) para elaborar predicciones precisas de la calidad de vídeo y ofrece una alternativa viable a las lentas y costosas evaluaciones subjetivas formales.

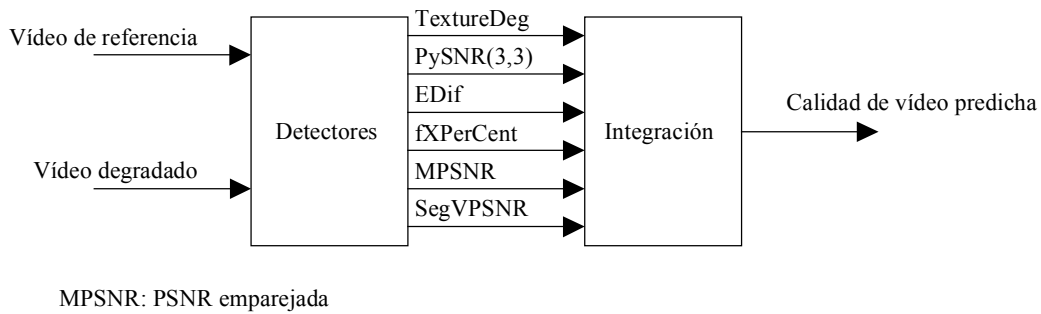
En las pruebas VQEG2 se utilizó una implementación del modelo en soporte lógico, y la calidad de funcionamiento resultante se presentó en un informe de la prueba<sup>1</sup>.

## 2 Modelo de BTFR

El algoritmo BTFR consiste en una detección seguida de una integración como se muestra en la Fig. 1. La detección supone el cálculo de un conjunto de parámetros perceptuales significativos del detector, a partir de secuencias de vídeo sin distorsionar (de referencia) y distorsionadas (degradadas). Estos parámetros se introducen a continuación en el integrador que produce una estimación de la calidad de vídeo percibida mediante la oportuna ponderación. La selección de detectores y factores de ponderación se basa en el conocimiento de las propiedades de enmascaramiento espacial y temporal del HVS y se determina mediante experimentos de calibración.

Este modelo soporta los tipos de vídeo de entrada 625 (720 × 576) entrelazado a 50 campos/s y 525 (720 × 486) entrelazado a 59,94 campos/s en el formato *YUV422*.

FIGURA 1  
Modelo de evaluación de la calidad de vídeo de referencia completa

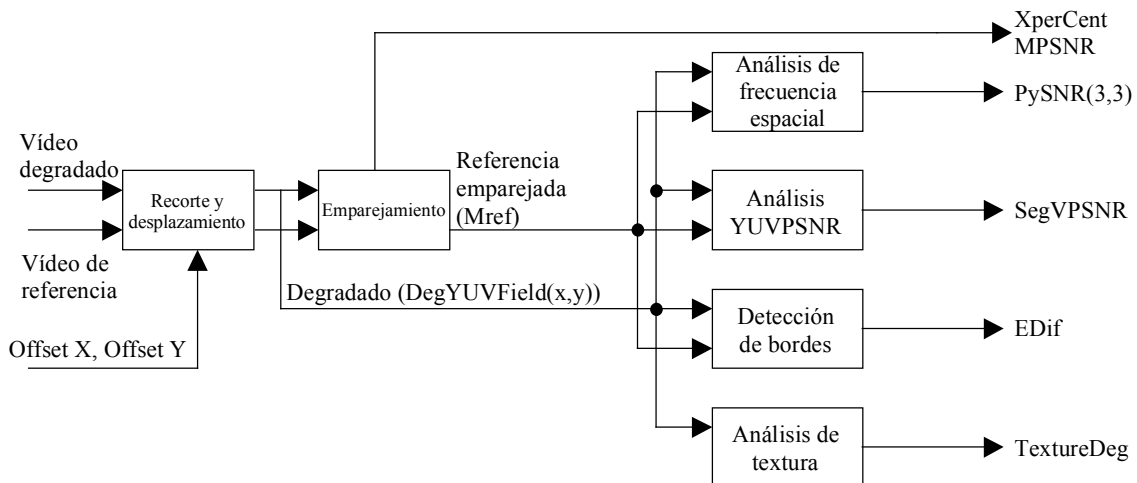


1683-01

### 3 Detectores

El módulo de detección del algoritmo BTFR calcula ciertas medidas espaciales, temporales y basadas en frecuencia a partir de las secuencias de entrada en formato *YUV*, como muestra la Fig. 2.

FIGURA 2  
Detección



1683-02

#### 3.1 Conversión de la entrada

En primer lugar, se convierten las secuencias de entrada del formato entrelazado *YUV422* al formato no entrelazado *YUV444* de bloque, de modo que cada campo sucesivo se representa mediante las matrices *RefY*, *RefU* y *RefV*:

$$RefY(x, y) \quad x = 0 \dots X - 1, \quad y = 0 \dots Y - 1 \quad (1)$$

$$RefU(x, y) \quad x = 0 \dots X - 1, \quad y = 0 \dots Y - 1 \quad (2)$$

$$RefV(x, y) \quad x = 0 \dots X - 1, \quad y = 0 \dots Y - 1 \quad (3)$$

siendo:

*X*: número de píxeles horizontales de un campo

*Y*: número de píxeles verticales.

Para una entrada *YUV422*, debe repetirse cada valor *U* y *V* para obtener las matrices (2) y (3) de resolución completa.

### 3.2 Recorte y desplazamiento

Esta rutina recorta con desplazamiento la secuencia de entrada degradada y recorta sin desplazamiento la secuencia de entrada de referencia. Los parámetros de desplazamiento  $XOffset$  y  $YOffset$  se determinan externamente y definen el número de píxeles horizontales y verticales en que la secuencia degradada está desplazada con respecto a la de referencia. El origen de la imagen se define en el ángulo superior izquierdo de la misma, siendo el incremento horizontal positivo hacia la derecha y el incremento vertical positivo hacia abajo con respecto a la imagen. El valor  $XOffset = 2$  indica que los campos degradados están desplazados a la derecha dos píxeles mientras que el valor  $YOffset = 2$  indica un desplazamiento hacia abajo de dos píxeles. Para un campo de entrada cuyos valores  $YUV$  estén almacenados en formato  $YUV444$  (véase el § 3.1) en las matrices  $InYField$ ,  $InUField$ , e  $InVField$ , la salida recortada y desplazada se calcula mediante las ecuaciones (4) a (20).

$$XStart = -XOffset \quad (4)$$

$$\text{si } (XStart < C_x) \quad \text{entonces } XStart = C_x \quad (5)$$

$$XEnd = X - 1 - XOffset \quad (6)$$

$$\text{si } (XEnd > X - C_x - 1) \quad \text{entonces } XEnd = X - C_x - 1 \quad (7)$$

$$YStart = -YOffset \quad (8)$$

$$\text{si } (YStart < C_y) \quad \text{entonces } YStart = C_y \quad (9)$$

$$YEnd = Y - 1 - YOffset \quad (10)$$

$$\text{si } (YEnd > Y - C_y - 1) \quad \text{entonces } YEnd = Y - C_y - 1 \quad (11)$$

$X$  e  $Y$  son las dimensiones horizontal y vertical del campo, respectivamente.  $C_x$  y  $C_y$  son el número de píxeles a recortar de izquierda a derecha y de arriba a abajo.

Para las secuencias de 625,

$$X = 720, \quad Y = 288, \quad C_x = 30, \quad C_y = 10 \quad (12)$$

Para las secuencias de 525,

$$X = 720, \quad Y = 243, \quad C_x = 30, \quad C_y = 10 \quad (13)$$

$Xstart$ ,  $Xend$ ,  $Ystart$  e  $Yend$  definen ahora la región que se copiará de cada campo. Los píxeles que queden fuera de esta región se inicializarán con arreglo a las ecuaciones (14) a (15), siendo  $YField$ ,  $UField$  y  $VField$  matrices de píxeles de salida  $XxY$  que contienen los valores de  $Y$ ,  $U$  y  $V$  respectivamente.

Las barras verticales a la izquierda y a la derecha del campo se inicializan con arreglo a las siguientes expresiones:

$$YField(x, y) = 0 \quad x = 0 \dots XStart - 1, XEnd + 1 \dots X - 1 \quad y = 0 \dots Y - 1 \quad (14)$$

$$UField(x, y) = VField(x, y) = 128 \quad x = 0 \dots XStart - 1, XEnd + 1 \dots X - 1 \quad y = 0 \dots Y - 1 \quad (15)$$

Las barras horizontales en la parte superior e inferior del campo se inicializan con arreglo a las siguientes expresiones:

$$YField(x, y) = 0 \quad x = XStart \dots XEnd, \quad y = 0 \dots YStart - 1, YEnd + 1 \dots Y - 1 \quad (16)$$

$$UField(x, y) = VField(x, y) = 128 \quad x = XStart \dots XEnd \quad y = 0 \dots YStart - 1, YEnd + 1 \dots Y - 1 \quad (17)$$

Por último, los valores de los píxeles se copian con arreglo a las siguientes expresiones:

$$YField(x, y) = InYField(x + XOffset, y + YOffset) \quad x = XStart...XEnd \quad y = YStart...YEnd \quad (18)$$

$$UField(x, y) = InUField(x + XOffset, y + YOffset) \quad x = XStart...XEnd \quad y = YStart...YEnd \quad (19)$$

$$VField(x, y) = InVField(x + XOffset, y + YOffset) \quad x = XStart...XEnd \quad y = YStart...YEnd \quad (20)$$

Para la entrada degradada, el recorte con desplazamiento produce las matrices de campos de salida  $DegYField$ ,  $DegUField$  y  $DegVField$ , mientras que el recorte sin desplazamiento para la secuencia de referencia produce  $RefYField$ ,  $RefUField$  y  $RefVfield$ . Estas matrices bidimensionales  $XxY$  se utilizan como entrada a las rutinas de detección descritas más adelante.

### 3.3 Emparejamiento

El proceso de emparejamiento genera señales para otros procedimientos de detección así como parámetros de detección para el procedimiento de integración. Las señales emparejadas se generan a partir de un proceso de búsqueda del emparejamiento óptimo para los bloques pequeños de cada campo degradado de una memoria intermedia de campos de referencia vecinos. Este proceso genera una secuencia, denominada referencia emparejada, que sustituye a la secuencia de referencia en algunos módulos de detección.

El análisis de emparejamiento se ejecuta en bloques de  $9 \times 9$  píxeles de las matrices de intensidad  $RefYField$  y  $DegYField$ . Añadiendo una dimensión numérica de campo a las matrices de intensidad, el píxel  $(Px, Py)$  del campo  $N$  de referencia puede representarse del siguiente modo:

$$Ref(N, Px, Py) = RefYField(Px, Py) \quad \text{del campo } N \quad (21)$$

Un bloque de  $9 \times 9$  píxeles centrado en el píxel  $(Px, Py)$  del  $N$ -ésimo campo puede representarse del siguiente modo:

$$BlockRef(N, Px, Py) = Ref(n, x, y) \quad x = Px - 4...Px + 4, \quad y = Py - 4...Py + 4 \quad (22)$$

$Deg(n, x, y)$  y  $BlockDeg(n, x, y)$  pueden definirse del mismo modo.

Para  $BlockDeg(N, Px, Py)$ , se calcula el error de adaptación mínimo,  $E(N, Px, Py)$ , buscando campos de referencia vecinos con arreglo a la siguiente expresión:

$$E(N, Px, Py) = \text{Min}((1/81) \sum_{j=-4}^4 \sum_{k=-4}^4 (Deg(N, Px + j, Py + k) - Ref(n, x + j, y + k))^2) \quad (23)$$

$$n = N - 4, \dots, N + 5$$

$$x = Px - 4, \dots, Px, \dots, Px + 4$$

$$y = Py - 4, \dots, Py, \dots, Py + 4$$

siendo  $N$  el índice del campo degradado que contiene el bloque degradado objeto de emparejamiento.



Si mediante la ecuación (23) se determina que la mejor pareja de  $BlockDeg(N, Px, Py)$  es  $BlockRef(n_m, x_m, y_m)$ , la matriz de referencia adaptada  $MRef$  se actualiza con arreglo a la siguiente expresión:

$$MRef(N, Px + j, Py + k) = Ref(n_m, x_m + j, y_m + k) \quad j = -4...4, k = -4...4 \quad (24)$$

El proceso de emparejamiento, consistente en buscar en primer lugar la mejor pareja para un bloque degradado y copiar a continuación el bloque resultante en la matriz de referencia emparejada, se repite para toda la región de análisis deseada. Esta región de análisis viene definida por los puntos centrales del bloque  $Px()$  y  $Py()$  del siguiente modo:

$$Px(h) = 16 + 8 \times h \quad h = 0...Qx - 1 \quad (25)$$

y

$$Py(v) = 16 + 8 \times v \quad v = 0...Qy - 1 \quad (26)$$

donde  $Qx$  y  $Qy$  definen el número de bloques de análisis horizontales y verticales.

Por consiguiente, el análisis de emparejamiento del  $N$ -ésimo campo produce una secuencia de referencia emparejada descrita por:

$$BlockMRef(N, Px(h), Py(v)) \quad h = 0...Qx - 1, \quad v = 0...Qy - 1 \quad (27)$$

y un conjunto de valores de error del mejor emparejamiento:

$$E(N, Px(h), Py(v)) \quad h = 0...Qx - 1, \quad v = 0...Qy - 1 \quad (28)$$

Puede definirse un conjunto de matrices de desplazamiento  $MatT$ ,  $MatX$  y  $MatY$  tal que:

$$BlockMRef(N, Px(h), Py(v)) = BlockRef(MatT(h, v), MatX(h, v), MatY(h, v)) \quad (29)$$

$$h = 0...Qx - 1, \quad v = 0...Qy - 1$$

Los parámetros de emparejamiento para secuencias de radiodifusión de 625 y 525 se presentan en el Cuadro 2.

## CUADRO 2

### Parámetros de búsqueda para el procedimiento de emparejamiento

Parámetro	625	525
$Qx$	87	87
$Qy$	33	28

La región de análisis definida por las ecuaciones (26) y (27) no ocupa todo el tamaño del campo. Por ello debe inicializarse  $MRef$  con arreglo a la ecuación (29) para poder utilizarla sin restricciones en cualquier sitio.

$$MRef(x, y) = 0 \quad x = 0...X - 1, \quad y = 0...Y - 1 \quad (30)$$

### 3.3.1 Estadísticas de emparejamiento

Se elaboran estadísticas de emparejamiento horizontal a partir del proceso de emparejamiento para utilizarlas en el proceso de integración. El mejor emparejamiento para cada bloque de análisis, determinado con arreglo a la ecuación (23), se utiliza en la construcción del histograma  $histX$  para cada campo con arreglo a la siguiente ecuación:

$$\begin{aligned} histX(MatX(h, v) - Px(h) + 4) &= histX(MatX(h, v) - Px(h) + 4) + 1 \\ h = 0 \dots Qx - 1, \quad v &= 0 \dots Qx - 1 \end{aligned} \quad (31)$$

donde la matriz  $histX$  se inicializa a cero para cada campo. A continuación se utiliza el histograma para determinar la medición  $fXPerCent$  de acuerdo con la ecuación:

$$fXPerCent = 100 \times \text{Max}(histX(i)) / \sum_{j=0}^8 histX(j) \quad i = 0 \dots 8 \quad (32)$$

Para cada campo, la medición  $fXPerCent$  es el porcentaje (%) de bloques emparejados que contribuye a la cresta del histograma de emparejamiento.

### 3.3.2 MPSNR

Se utiliza el error mínimo,  $E()$ , de cada bloque emparejado para calcular la SNR emparejada con arreglo a las siguientes ecuaciones:

$$\text{si } \left( \sum_{h=0}^{Qx-1} \sum_{v=0}^{Qy-1} E(N, Px(h), Py(v)) \right) > 0 \quad \text{entonces} \quad (33)$$

$$MPSNR = 10 \log_{10} \left( Qx \times Qy \times 255^2 / \sum_{h=0}^{Qx-1} \sum_{v=0}^{Qy-1} E(N, Px(h), Py(v)) \right)$$

$$\text{si } \left( \sum_{h=0}^{Qx-1} \sum_{v=0}^{Qy-1} E(N, Px(h), Py(v)) \right) = 0 \quad \text{entonces} \quad MPSNR = 10 \log_{10}(255^2) \quad (34)$$

### 3.3.3 Vectores de emparejamiento

Los vectores horizontal, vertical y de retardo se guardan para ser utilizados más adelante, con arreglo a las siguientes expresiones:

$$SyncT(h, v) = MatT(h, v) - N \quad h = 0 \dots Qx - 1, \quad v = 0 \dots Qy - 1 \quad (35)$$

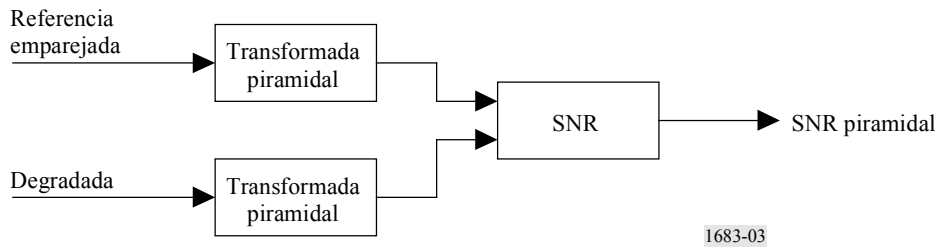
$$SyncX(h, v) = MatX(h, v) - Px(h) \quad h = 0 \dots Qx - 1, \quad v = 0 \dots Qy - 1 \quad (36)$$

$$SyncY(h, v) = MatY(h, v) - Py(h) \quad h = 0 \dots Qx - 1, \quad v = 0 \dots Qy - 1 \quad (37)$$

### 3.4 Análisis espacial de frecuencia

El detector espacial de frecuencia se basa en una transformación «piramidal» de las secuencias de referencia degradada y emparejada. En primer lugar se transforma cada secuencia generando las matrices piramidales de referencia y degradada. A continuación se calculan las diferencias entre las matrices piramidales midiendo los errores cuadráticos medios, obteniendo los resultados como SNR piramidal.

FIGURA 3  
Análisis espacial de frecuencia



#### 3.4.1 La transformada piramidal

En primer lugar, el campo de entrada,  $F$ , se copia a una matriz piramidal,  $P$ , con arreglo a la siguiente expresión:

$$P(x, y) = F(x, y) \quad x = 0 \dots X - 1, \quad y = 0 \dots Y - 1 \quad (38)$$

A continuación se actualiza esta matriz piramidal en tres etapas ( $stage = 0 \dots 2$ ) de análisis vertical y horizontal. El análisis horizontal  $Hpy(stage)$  viene definido por las ecuaciones (39) a (43).

Primero se hace una copia temporal de toda la matriz piramidal:

$$PTemp(x, y) = P(x, y) \quad x = 0 \dots X - 1, \quad y = 0 \dots Y - 1 \quad (39)$$

A continuación se calculan los límites  $x$  e  $y$  de acuerdo con:

$$Tx = X/2^{(stage+1)} \quad (40)$$

$$Ty = Y/2^{stage} \quad (41)$$

A continuación se utilizan los promedios y las diferencias de los pares de elementos horizontales de la matriz temporal para actualizar la matriz piramidal con arreglo a las siguientes ecuaciones:

$$P(x, y) = 0,5 (PTemp(2x, y) + PTemp(2x + 1, y)) \quad x = 0 \dots Tx - 1, \quad y = 0 \dots Ty - 1 \quad (42)$$

$$P(x + Tx, y) = PTemp(2x, y) - PTemp(2x + 1, y) \quad x = 0 \dots Tx - 1 \quad y = 0 \dots Ty - 1 \quad (43)$$

El análisis vertical  $Vpy(stage)$  viene definido por las ecuaciones (44) a (48).

$$PTemp(x, y) = P(x, y) \quad x = 0 \dots X - 1, \quad y = 0 \dots Y - 1 \quad (44)$$

$$Tx = X/2^{stage} \quad (45)$$

$$Ty = Y/2^{(stage+1)} \quad (46)$$

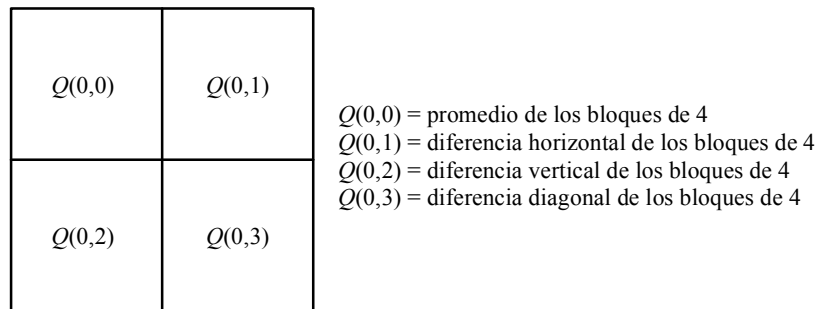
A continuación se utilizan los promedios y diferencias de los pares de elementos verticales de la matriz temporal para actualizar la matriz piramidal con arreglo a las siguientes expresiones:

$$P(x, y) = 0,5 (PTemp(x, 2y) + PTemp(x, 2y + 1)) \quad x = 0 \dots Tx - 1, \quad y = 0 \dots Ty - 1 \quad (47)$$

$$P(x, y + Ty) = PTemp(x, 2y) - PTemp(x, 2y + 1) \quad x = 0 \dots Tx - 1 \quad y = 0 \dots Ty - 1 \quad (48)$$

En la Etapa 0, el análisis horizontal  $Hpy(0)$  seguido del análisis vertical  $Vpy(0)$  actualiza toda la matriz piramidal con los 4 cuadrantes  $Q(stage, 0 \dots 3)$  construidos del siguiente modo:

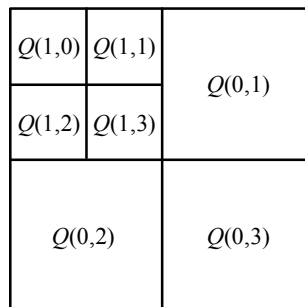
FIGURA 4  
Cuadrante de salida del análisis de la etapa 0



1683-04

A continuación se efectúa el análisis de la Etapa 1 sobre  $Q(0,0)$  y los resultados  $Q(1,0 \dots 3)$  generados se almacenan en la pirámide con arreglo a:

FIGURA 5  
Cuadrante de salida del análisis de la etapa 1



1683-05

En el análisis de la Etapa 2 se procesa  $Q(1,0)$  y se sustituye por  $Q(2,0 \dots 3)$ .

Tras la tercera etapa del análisis, la matriz piramidal resultante tendrá un total de 10 bloques de resultados. Tres bloques  $Q(0,1 \dots 3)$  procederán del análisis de  $2 \times 2$  píxeles de la Etapa 0, tres  $Q(1,1 \dots 3)$  del análisis  $4 \times 4$  de las Etapas 1 y 4  $Q(2,0 \dots 3)$  del análisis  $8 \times 8$  de la Etapa 2.

Del análisis de tres etapas de las secuencias emparejadas de referencia y degradada se obtienen las matrices piramidales  $Pref$  y  $Pdeg$ . Las diferencias entre estas matrices se miden a continuación en el módulo SNR piramidal.

### 3.4.2 SNR piramidal

La medida del error cuadrático entre las matrices piramidales de referencia y degradada se determina en los cuadrantes 1 a 3 de las Etapas 0 a 2 con arreglo a la siguiente ecuación:

$$E(s, q) = (1/XY^2) \sum_{x=x1(s, q)}^{x2(s, q)-1} \sum_{y=y1(s, q)}^{y2(s, q)-1} (Pref(x, y) - Pdeg(x, y))^2 \quad s = 0...2 \quad q = 1...3 \quad (49)$$

donde  $x1$ ,  $x2$ ,  $y1$  e  $y2$  definen los límites horizontales y verticales de los cuadrantes de las matrices piramidales y se calculan con arreglo a las siguientes ecuaciones:

$$x1(s, 1) = X/2^{(s+1)} \quad x2(s, 1) = 2 \times x1(s, 1) \quad y1(s, 1) = 0 \quad y2(s, 1) = Y/2^{(s+1)} \quad (50)$$

$$x1(s, 2) = 0 \quad x2(s, 2) = X/2^{(s+1)} \quad y1(s, 2) = Y/2^{(s+1)} \quad y2(s, 2) = 2 \times y1(s, 2) \quad (51)$$

$$x1(s, 3) = X/2^{(s+1)} \quad x2(s, 3) = 2 \times x1(s, 3) \quad y1(s, 3) = Y/2^{(s+1)} \quad y2(s, 3) = 2 \times y1(s, 3) \quad (52)$$

A continuación se utilizan los resultados de la expresión (49) para medir la PSNR de cada uno de los cuadrantes de cada campo con arreglo a la siguiente ecuación:

$$\begin{aligned} \text{si } (E > 0, 0) \quad & PySNR(s, q) = 10 \log_{10}(255^2/E(s, q)) \\ \text{entonces} \quad & SNR = 10 \log_{10}(255^2 \times XY^2) \end{aligned} \quad (53)$$

siendo el número de Etapas  $s = 0...2$  y el número de cuadrantes de cada Etapa  $q = 1...3$ .

### 3.5 Análisis de textura

La textura de la secuencia degradada se mide registrando el número de puntos de inflexión de la señal de intensidad a lo largo de las líneas horizontales de imagen. Esto puede calcularse mediante las ecuaciones (54) a (59).

Para cada campo se inicializa primero un contador de puntos de inflexión con arreglo a la ecuación (54).

$$sum = 0 \quad (54)$$

A continuación, se procesa cada línea,  $y = 0...Y - 1$ , para  $x = 0...X - 2$  con arreglo a las siguientes ecuaciones:

$$last\_pos = 0, \quad last\_neg = 0 \quad (55)$$

$$dif(x) = P(x, y) - P(x+1, y) \quad (56)$$

$$\text{si } ((dif(x) < 0) \text{ y } (last\_neg < last\_pos)) sum = sum + 1 \quad (57)$$

$$\text{si } ((dif(x) > 0) \text{ y } (last\_neg > last\_pos)) sum = sum + 1 \quad (58)$$

$$\text{si } (dif(x) > 0) last\_pos = x \quad (59)$$

$$\text{si } (dif(x) < 0) last\_neg = x \quad (60)$$

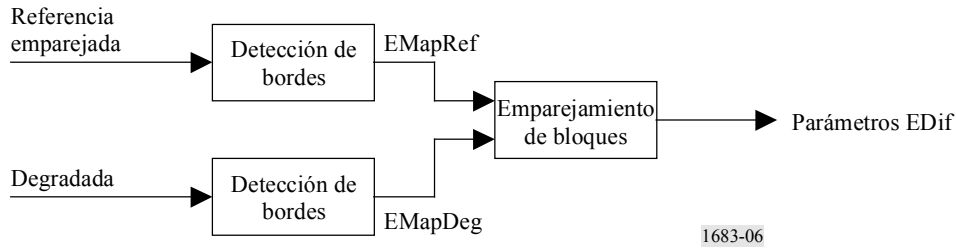
Una vez procesadas todas las líneas de un campo, el contador  $sum$  contendrá el número de puntos de inflexión de la señal de intensidad horizontal, que se utiliza para calcular un parámetro de textura para cada campo de la siguiente forma:

$$TextureDeg = sum \times 100/XY \quad (61)$$

### 3.6 Análisis de los bordes

Cada uno de los campos de las secuencias de referencia degradada y emparejada se somete por separado a una rutina de detección de bordes para generar los correspondientes mapas de bordes y compararlos mediante un procedimiento de emparejamiento de bloques para generar los parámetros de detección.

FIGURA 6  
Análisis de los bordes



#### 3.6.1 Detección de los bordes

Se ha utilizado un detector de bordes Canny [Canny, 1986] para determinar los mapas de bordes, aunque pueden utilizarse otras técnicas similares de detección de bordes. Los mapas de bordes resultantes,  $EMapRef$  y  $EMapDeg$ , son mapas de píxeles en los que los bordes se indican por 1 y su ausencia por 0.

Para un borde detectado en el píxel  $(x, y)$ :

$$EMap(x, y) = 1 \quad x = 0 \dots X - 1, \quad y = 0 \dots Y - 1 \quad (62)$$

Si no se ha detectado borde alguno en el píxel  $(x, y)$ :

$$EMap(x, y) = 0 \quad x = 0 \dots X - 1, \quad y = 0 \dots Y - 1 \quad (63)$$

#### 3.6.2 Diferenciación de los bordes

El procedimiento de diferenciación de los bordes mide las diferencias entre los mapas de bordes de los campos de referencia degradado y emparejado correspondientes. El análisis se efectúa en bloques no solapados de  $N \times M$  píxeles con arreglo a las ecuaciones (64) a (68).

En primer lugar se mide el número de píxeles marcados con borde en cada bloque de análisis, donde  $Bh$  y  $Bv$  definen el número de bloques no solapados a analizar en las direcciones horizontal y vertical y  $X1$  e  $Y1$  definen los desplazamientos del análisis con respecto al borde del campo.

$$Bref(x, y) = \sum_{i=i1}^{i2} \sum_{j=j1}^{j2} EMapRef(Nx + X1 + i, My + Y1 + j) \quad x = 0 \dots Bh - 1, y = 0 \dots Bv - 1 \quad (64)$$

$$BDeg(x, y) = \sum_{i=i1}^{i2} \sum_{j=j1}^{j2} EMapDeg(Nx + X1 + i, My + Y1 + j) \quad x = 0 \dots Bh - 1, y = 0 \dots Bv - 1 \quad (65)$$

Los límites de la suma se determinan de acuerdo con las siguientes ecuaciones:

$$i1 = -(N \text{ div } 2) \quad i2 = (N - 1) \text{ div } 2 \quad (66)$$

$$j1 = -(M \text{ div } 2) \quad j2 = (M - 1) \text{ div } 2 \quad (67)$$

donde el operador «div» representa la división entera.

A continuación se calcula la medida de las diferencias en todo el campo mediante la siguiente ecuación:

$$EDif = (1/(N M Bh Bv)) \left( \sum_{x=0}^{Bh-1} \sum_{y=0}^{Bv-1} (BRef(x,y) - BDeg(x,y))^Q \right)^{1/Q} \quad (68)$$

Para los campos de  $720 \times 288$  píxeles en el vídeo de radiodifusión de 625:

$$N = 4, \quad X1 = 6, \quad Bh = 178 \quad M = 4, \quad Y1 = 10, \quad Bv = 69, \quad Q = 3 \quad (69)$$

Para los campos de  $720 \times 243$  píxeles en el vídeo de radiodifusión de 525:

$$N = 4, \quad X1 = 6, \quad Bh = 178 \quad M = 4, \quad Y1 = 10, \quad Bv = 58, \quad Q = 3 \quad (70)$$

### 3.7 Análisis de la MPSNR

Se calcula la relación señal-ruido emparejada para los valores del píxel  $V$  utilizando los vectores de emparejamiento definidos por las ecuaciones (35) a (37). Para cada conjunto de vectores emparejados se calcula la medida del error,  $VE$ , mediante la ecuación:

$$VE(h,v) = (1/81) \sum_{i=-4}^4 \sum_{j=-4}^4 (DegV(N, Px(h)+i, Py(h)+j) - RefVField(N + SyncT(h,v), Px(h) + SyncX(h,v) + i, Py(v) + SyncY(h,v) + j))^2 \quad (71)$$

A continuación se calcula la medida de la PSNR segmentaria del campo mediante la ecuación (72):

$$SegVPSNR = (1/Qx Qy) \sum_{h=0}^{Qx-1} \sum_{v=0}^{Qy-1} 10 \log_{10}(255^2 / (VE(h,v) + 1)) \quad (72)$$

## 4 Integración

El procedimiento de integración requiere en primer lugar el promedio en el tiempo de los parámetros de detección campo a campo, con arreglo a la ecuación (73):

$$AvD(k) = (1/N) \sum_{n=0}^{N-1} D(k,n) \quad k = 0...5 \quad (73)$$

siendo:

$N$ : número total de campos de las secuencias probadas

$D(k,n)$ : parámetro de detección  $k$  del campo  $n$ .

A continuación, se combinan los parámetros de detección promediados,  $AvD(k)$ , para proporcionar una predicción de la nota de calidad, PDMOS, correspondiente a la secuencia del campo  $N$  de acuerdo con la ecuación (74):

$$PDMOS = Offset + \sum_{k=0}^5 AvD(k) \times W(k) \quad (74)$$

Los Cuadros 3 y 4 muestran los parámetros del integrador para las secuencias de 625 y 525 respectivamente.

CUADRO 3

**Parámetros de integración para el vídeo de radiodifusión de 625**

$K$	Nombre del parámetro	$W$
0	<i>TextureDeg</i>	-0,68
1	<i>PySNR(3,3)</i>	-0,57
2	<i>EDif</i>	+58 913,294
3	<i>fXPerCent</i>	-0,208
4	<i>MPSNR</i>	-0,928
5	<i>SegVPSNR</i>	-1,529
Offset	+176,486	
$N$	400	

CUADRO 4

**Parámetros de integración para el vídeo de radiodifusión de 525**

$K$	Nombre del parámetro	$W$
0	<i>TextureDeg</i>	+0,043
1	<i>PySNR(3,3)</i>	-2,118
2	<i>EDif</i>	+60 865,164
3	<i>fXPerCent</i>	-0,361
4	<i>MPSNR</i>	+1,104
5	<i>SegVPSNR</i>	-1,264
Offset	+260,773	
$N$	480	

**5 Registro**

Para que el modelo FR funcione con eficacia es necesaria tanto la alineación espacial como la temporal. El modelo incorpora intrínsecamente la alineación y puede acomodar desplazamientos espaciales entre la secuencia degradada y la de referencia de  $\pm 4$  píxeles y desplazamientos temporales de  $\pm 4$  campos. Los desplazamientos espaciales y temporales que superen estos límites no son soportados por este modelo siendo necesario un módulo de registro independiente para garantizar la correcta alineación de los ficheros de referencia y degradado.

**6 Referencias Bibliográficas**

CANNY, J. [1986] A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 8(6), p. 679-698.



## Anexo 2a

## CUADRO 5

## Datos subjetivos y objetivos de 525

Nombre del fichero	Secuencia fuente (SRC)	Circuito de referencia hipotético (HRC)	Índice subjetivo medio bruto	Índice predicho por el modelo a partir de los datos brutos	Índice subjetivo medio ajustado a escala	Índice predicho por el modelo a partir de datos ajustados a escala
V2src01_hrc01_525.yuv	1	1	-38,30757576	44,945049	0,5402368	0,69526
V2src01_hrc02_525.yuv	1	2	-39,56212121	38,646271	0,5483205	0,58989
V2src01_hrc03_525.yuv	1	3	-25,9469697	32,855755	0,4024097	0,50419
V2src01_hrc04_525.yuv	1	4	-17,24090909	21,062775	0,3063528	0,36089
V2src02_hrc01_525.yuv	2	1	-35,23636364	31,260744	0,5025558	0,48242
V2src02_hrc02_525.yuv	2	2	-18,01818182	18,732758	0,3113346	0,33715
V2src02_hrc03_525.yuv	2	3	-6,284848485	8,914509	0,1881739	0,25161
V2src02_hrc04_525.yuv	2	4	-6,983333333	4,16663	0,1907347	0,21776
V2src03_hrc01_525.yuv	3	1	-31,96515152	22,348713	0,4682724	0,37461
V2src03_hrc02_525.yuv	3	2	-17,47727273	10,44728	0,3088831	0,26352
V2src03_hrc03_525.yuv	3	3	-1,104545455	2,494911	0,1300389	0,20688
V2src03_hrc04_525.yuv	3	4	-1,171212121	0	0,1293293	0,19158
V2src04_hrc05_525.yuv	4	5	-50,64090909	40,82526	0,6742005	0,6249
V2src04_hrc06_525.yuv	4	6	-28,05454545	32,552322	0,4250873	0,49999
V2src04_hrc07_525.yuv	4	7	-23,87575758	25,286598	0,3762656	0,40764
V2src04_hrc08_525.yuv	4	8	-16,60757576	19,86405	0,2972294	0,3485
V2src05_hrc05_525.yuv	5	5	-31,86969697	30,812616	0,4682559	0,47645
V2src05_hrc06_525.yuv	5	6	-18,56515152	21,413895	0,3203024	0,3646
V2src05_hrc07_525.yuv	5	7	-8,154545455	15,446437	0,2071702	0,306
V2src05_hrc08_525.yuv	5	8	-4,006060606	10,836051	0,1652752	0,26662
V2src06_hrc05_525.yuv	6	5	-41,63181818	37,342789	0,5690291	0,56967
V2src06_hrc06_525.yuv	6	6	-29,48787879	26,660055	0,4370961	0,42391
V2src06_hrc07_525.yuv	6	7	-22,25909091	20,878248	0,3591788	0,35896
V2src06_hrc08_525.yuv	6	8	-12,03181818	16,896168	0,2482169	0,31941
V2src07_hrc05_525.yuv	7	5	-23,89545455	19,086998	0,3796362	0,34067
V2src07_hrc06_525.yuv	7	6	-10,15606061	10,69402	0,2276934	0,26548
V2src07_hrc07_525.yuv	7	7	-4,240909091	4,896546	0,1644409	0,22267
V2src07_hrc08_525.yuv	7	8	-5,98030303	1,555055	0,1819566	0,20099
V2src08_hrc09_525.yuv	8	9	-76,2	52,094177	0,9513387	0,83024
V2src08_hrc10_525.yuv	8	10	-61,34545455	47,395226	0,789748	0,7397
V2src08_hrc11_525.yuv	8	11	-66,02575758	52,457584	0,8405916	0,83753
V2src08_hrc12_525.yuv	8	12	-37,20454545	37,931854	0,5221555	0,57874
V2src08_hrc13_525.yuv	8	13	-31,23030303	30,95985	0,4572049	0,4784

## CUADRO 5

## Datos subjetivos y objetivos de 525

Nombre del fichero	Secuencia fuente (SRC)	Circuito de referencia hipotético (HRC)	Índice subjetivo medio bruto	Índice predicho por el modelo a partir de los datos brutos	Índice subjetivo medio ajustado a escala	Índice predicho por el modelo a partir de datos ajustados a escala
V2src08_hrc14_525.yuv	8	14	-31,26818182	33,293602	0,4614104	0,51031
V2src09_hrc09_525.yuv	9	9	-64,42878788	54,414772	0,8262912	0,87746
V2src09_hrc10_525.yuv	9	10	-49,92878788	36,080425	0,660339	0,55061
V2src09_hrc11_525.yuv	9	11	-53,73181818	46,338791	0,7100111	0,72031
V2src09_hrc12_525.yuv	9	12	-34,36969697	23,21393	0,4921708	0,38409
V2src09_hrc13_525.yuv	9	13	-22,85454545	16,955978	0,3656559	0,31998
V2src09_hrc14_525.yuv	9	14	-16,41666667	13,694396	0,2960957	0,29046
V2src10_hrc09_525.yuv	10	9	-72,11212121	48,179104	0,9084171	0,75433
V2src10_hrc10_525.yuv	10	10	-43,11666667	30,703861	0,5908784	0,475
V2src10_hrc11_525.yuv	10	11	-56,11969697	52,63887	0,7302376	0,84118
V2src10_hrc12_525.yuv	10	12	-19,55909091	21,95225	0,3345703	0,37033
V2src10_hrc13_525.yuv	10	13	-12,34393939	16,23988	0,2565459	0,31328
V2src10_hrc14_525.yuv	10	14	-16,05	23,201355	0,2953144	0,38395
V2src11_hrc09_525.yuv	11	9	-50,40454545	36,394535	0,6675853	0,55531
V2src11_hrc10_525.yuv	11	10	-54,26212121	37,812542	0,7054929	0,5769
V2src11_hrc11_525.yuv	11	11	-41,73636364	44,128036	0,5761193	0,68087
V2src11_hrc12_525.yuv	11	12	-19,03939394	14,619688	0,32761	0,29857
V2src11_hrc13_525.yuv	11	13	-17,72121212	14,12041	0,310495	0,29417
V2src11_hrc14_525.yuv	11	14	-19,4969697	14,927424	0,331051	0,30132
V2src12_hrc09_525.yuv	12	9	-61,35	40,051254	0,7883371	0,61229
V2src12_hrc10_525.yuv	12	10	-46,84545455	31,128973	0,6295301	0,48066
V2src12_hrc11_525.yuv	12	11	-51,80151515	41,77285	0,6809288	0,6406
V2src12_hrc12_525.yuv	12	12	-22,51969697	20,868282	0,3651402	0,35886
V2src12_hrc13_525.yuv	12	13	-14,17878788	15,040992	0,2714356	0,30234
V2src12_hrc14_525.yuv	12	14	-14,6030303	13,521517	0,2782449	0,28896
V2src13_hrc09_525.yuv	13	9	-55,25	38,691498	0,7211194	0,5906
V2src13_hrc10_525.yuv	13	10	-39,55	33,054504	0,5545722	0,50696
V2src13_hrc11_525.yuv	13	11	-40,03939394	45,9454	0,5525494	0,71318
V2src13_hrc12_525.yuv	13	12	-14	16,631002	0,2708744	0,31692
V2src13_hrc13_525.yuv	13	13	-14,33181818	15,113959	0,27549	0,30299
V2src13_hrc14_525.yuv	13	14	-14,31969697	16,611286	0,2733771	0,31674

## CUADRO 6

## Datos subjetivos y objetivos de 625

Nombre del fichero	SRC	HRC	Índice subjetivo medio bruto	Índice predicho por el modelo a partir de los datos brutos	Índice subjetivo medio ajustado a escala	Índice predicho por el modelo a partir de datos ajustados a escala
V2src1_hrc2_625.yuv	1	2	38,85185185	31,764214	0,59461	0,47326
V2src1_hrc3_625.yuv	1	3	42,07407407	21,868561	0,64436	0,36062
V2src1_hrc4_625.yuv	1	4	23,77777778	12,195552	0,40804	0,27239
V2src1_hrc6_625.yuv	1	6	18,14814815	9,169512	0,34109	0,24887
V2src1_hrc8_625.yuv	1	8	12,92592593	6,738072	0,2677	0,23128
V2src1_hrc10_625.yuv	1	10	11,88888889	2,553883	0,26878	0,20356
V2src2_hrc2_625.yuv	2	2	33,51851852	31,492788	0,54173	0,46985
V2src2_hrc3_625.yuv	2	3	46,48148148	31,1313	0,70995	0,46535
V2src2_hrc4_625.yuv	2	4	13,33333333	20,241726	0,27443	0,34432
V2src2_hrc6_625.yuv	2	6	8,814814815	17,39045	0,22715	0,31721
V2src2_hrc8_625.yuv	2	8	7,074074074	14,914576	0,21133	0,29513
V2src2_hrc10_625.yuv	2	10	3,407407407	7,352309	0,16647	0,23562
V2src3_hrc2_625.yuv	3	2	48,07407407	38,852715	0,73314	0,56845
V2src3_hrc3_625.yuv	3	3	50,66666667	38,244621	0,76167	0,55982
V2src3_hrc4_625.yuv	3	4	32,11111111	27,733229	0,49848	0,42454
V2src3_hrc6_625.yuv	3	6	22,33333333	24,80323	0,38613	0,39159
V2src3_hrc8_625.yuv	3	8	16,33333333	23,296747	0,34574	0,37544
V2src3_hrc10_625.yuv	3	10	11,96296296	16,33028	0,26701	0,30759
V2src4_hrc2_625.yuv	4	2	36,14814815	42,041592	0,58528	0,61514
V2src4_hrc3_625.yuv	4	3	55,03703704	49,283836	0,90446	0,72942
V2src4_hrc4_625.yuv	4	4	39,7037037	38,322186	0,62361	0,56091
V2src4_hrc6_625.yuv	4	6	38,03703704	36,863457	0,61143	0,54053
V2src4_hrc8_625.yuv	4	8	24,40740741	32,46579	0,43329	0,48214
V2src4_hrc10_625.yuv	4	10	12,88888889	25,918123	0,26548	0,40388
V2src5_hrc2_625.yuv	5	2	38,62962963	38,95779	0,61973	0,56995
V2src5_hrc3_625.yuv	5	3	44,18518519	40,076313	0,68987	0,58609
V2src5_hrc4_625.yuv	5	4	24,66666667	23,166002	0,41648	0,37406
V2src5_hrc6_625.yuv	5	6	23,62962963	20,592213	0,4218	0,34778
V2src5_hrc8_625.yuv	5	8	12,40740741	13,763152	0,27543	0,28531
V2src5_hrc10_625.yuv	5	10	7,37037037	8,418313	0,2022	0,24332
V2src6_hrc2_625.yuv	6	2	22,48148148	33,810165	0,38852	0,49949
V2src6_hrc3_625.yuv	6	3	27,07407407	25,004984	0,44457	0,39379
V2src6_hrc4_625.yuv	6	4	13,18518519	20,889347	0,27983	0,35074
V2src6_hrc6_625.yuv	6	6	14,44444444	17,418222	0,28106	0,31747

## CUADRO 6

## Datos subjetivos y objetivos de 625

Nombre del fichero	SRC	HRC	Índice subjetivo medio bruto	Índice predicho por el modelo a partir de los datos brutos	Índice subjetivo medio ajustado a escala	Índice predicho por el modelo a partir de datos ajustados a escala
V2src6_hrc8_625.yuv	6	8	8,740740741	15,486559	0,23726	0,30011
V2src6_hrc10_625.yuv	6	10	5,518518519	11,509192	0,17793	0,2669
V2src7_hrc4_625.yuv	7	4	39,25925926	45,231079	0,59953	0,66412
V2src7_hrc6_625.yuv	7	6	33,85185185	43,131519	0,55093	0,63163
V2src7_hrc9_625.yuv	7	9	27,07407407	39,506535	0,45163	0,57784
V2src7_hrc10_625.yuv	7	10	19,25925926	34,418381	0,35617	0,50749
V2src8_hrc4_625.yuv	8	4	15,85185185	40,408993	0,32528	0,59095
V2src8_hrc6_625.yuv	8	6	17,03703704	38,552574	0,32727	0,56418
V2src8_hrc9_625.yuv	8	9	14,85185185	35,577034	0,30303	0,52297
V2src8_hrc10_625.yuv	8	10	11,48148148	30,278536	0,26366	0,45484
V2src9_hrc4_625.yuv	9	4	28,96296296	30,515778	0,47656	0,45775
V2src9_hrc6_625.yuv	9	6	30,51851852	26,971027	0,49924	0,41577
V2src9_hrc9_625.yuv	9	9	19,66666667	23,351355	0,39101	0,37601
V2src9_hrc10_625.yuv	9	10	20,92592593	17,856861	0,37122	0,32152
V2src10_hrc4_625.yuv	10	4	40,33333333	43,640377	0,70492	0,63942
V2src10_hrc6_625.yuv	10	6	37,33333333	40,552502	0,58218	0,59305
V2src10_hrc9_625.yuv	10	9	30,92592593	36,747391	0,49711	0,53893
V2src10_hrc10_625.yuv	10	10	21,2962963	30,161013	0,37854	0,45341
V2src11_hrc1_625.yuv	11	1	50,25925926	55,909908	0,79919	0,84263
V2src11_hrc5_625.yuv	11	5	35,51851852	44,049999	0,59256	0,64572
V2src11_hrc7_625.yuv	11	7	18,7037037	26,877754	0,34337	0,4147
V2src11_hrc10_625.yuv	11	10	15,07407407	23,420477	0,30567	0,37674
V2src12_hrc1_625.yuv	12	1	36,33333333	43,837097	0,61418	0,64244
V2src12_hrc5_625.yuv	12	5	38,44444444	40,349903	0,6661	0,59008
V2src12_hrc7_625.yuv	12	7	31,11111111	37,254383	0,53242	0,54594
V2src12_hrc10_625.yuv	12	10	26,14814815	28,953564	0,44737	0,43887
V2src13_hrc1_625.yuv	13	1	43,7037037	38,333649	0,74225	0,56108
V2src13_hrc5_625.yuv	13	5	43,2962963	34,290554	0,66799	0,5058
V2src13_hrc7_625.yuv	13	7	25,2962963	26,990025	0,42065	0,41598
V2src13_hrc10_625.yuv	13	10	15,88888889	20,181463	0,33381	0,34373

**Anexo 3****Modelo 2****ÍNDICE**

	<i>Página</i>
1	Introducción ..... 21
2	Medición objetiva de la calidad de vídeo a partir de la degradación de los bordes ..... 22
2.1	PSNR de los bordes (EPSNR) ..... 22
2.2	Ajustes posteriores ..... 29
2.2.1	De énfasis de EPSNR ..... 29
2.2.2	Consideración de los bordes borrosos ..... 29
2.2.3	Ajuste ..... 30
2.3	Precisión de registro ..... 30
2.4	Diagrama de bloques del modelo ..... 30
3	Datos objetivos ..... 30
4	Conclusión ..... 30
5	Referencias Bibliográficas ..... 30

**1 Introducción**

La evaluación de la calidad de vídeo se ha realizado tradicionalmente con evaluadores que determinan subjetivamente la calidad de vídeo. La evaluación puede llevarse a cabo con vídeos de referencia o sin ellos. En la evaluación con referencias, los evaluadores contemplan dos vídeos: el vídeo de referencia (fuente) y el vídeo procesado que ha de compararse con el vídeo fuente. Tras comparar los dos vídeos, los evaluadores otorgan a los vídeos notas subjetivas. Por este motivo se le suele denominar prueba subjetiva de la calidad de vídeo. Aunque se considera que la prueba subjetiva es el método más preciso debido a que refleja la percepción humana, tiene varias limitaciones. En primer lugar, requiere varios evaluadores, motivo por el cual es lenta y costosa. Además, no puede llevarse a cabo en tiempo real. Por consiguiente, el desarrollo de métodos objetivos para la medición de la calidad de vídeo siempre ha suscitado un gran interés. Uno de los

requisitos más importantes que debe satisfacer cualquier método objetivo de medición de la calidad de vídeo es proporcionar resultados de la calidad de funcionamiento coherentes para una amplia gama de secuencias de vídeo no utilizadas en la etapa de diseño. A tal efecto se ha desarrollado un modelo fácil de implementar, suficientemente rápido para las implementaciones en tiempo real y robusto frente a una amplia gama de degradaciones de vídeo. Este modelo es fruto de la colaboración entre la Yonsei University, SK Telecom y el Radio Research Laboratory, República de Corea.

## 2 Medición objetiva de la calidad de vídeo a partir de la degradación de los bordes

### 2.1 PSNR de los bordes (EPSNR)

El modelo para la medición objetiva de la calidad de vídeo es de referencia completa. O sea, se supone que se dispone de un vídeo de referencia. Al analizar la percepción por el hombre de la calidad de vídeo, se observa que el sistema visual humano es sensible a la degradación en torno a los bordes. Dicho de otro modo, cuando las zonas de los bordes de un vídeo son borrosas, los evaluadores tienden a otorgar puntuaciones menores al vídeo aunque el error cuadrático medio global sea pequeño. Se observa además que los algoritmos de compresión de vídeo tienden a producir más artefactos en torno a las zonas de los bordes. De acuerdo con esta observación, el modelo proporciona un método de medición objetiva de la calidad de vídeo que mide la degradación en torno a los bordes. En este modelo, se aplica en primer lugar un algoritmo de detección de bordes a la secuencia de vídeo fuente para localizar las zonas de los bordes. A continuación, se mide la degradación de dichas zonas de bordes calculando el error cuadrático medio. A partir de dicho error se calcula la EPSNR del borde, que se utiliza como métrica de la calidad de vídeo tras el postprocesamiento.

En este modelo, es necesario aplicar primero un algoritmo de detección de bordes para localizar las zonas de los bordes. Se puede utilizar cualquier algoritmo de detección de bordes, aunque puede haber diferencias de menor importancia en los resultados. Por ejemplo, se puede utilizar cualquier operador gradiente para localizar las zonas de los bordes. Se han propuesto varios operadores gradiente. En muchos algoritmos de detección de bordes, la imagen de gradiente horizontal  $g_{horizontal}(m,n)$  y la imagen de gradiente vertical  $g_{vertical}(m,n)$  se calculan previamente mediante operadores gradiente. A continuación, puede calcularse la imagen de gradiente de magnitud  $g(m,n)$  del siguiente modo:

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

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

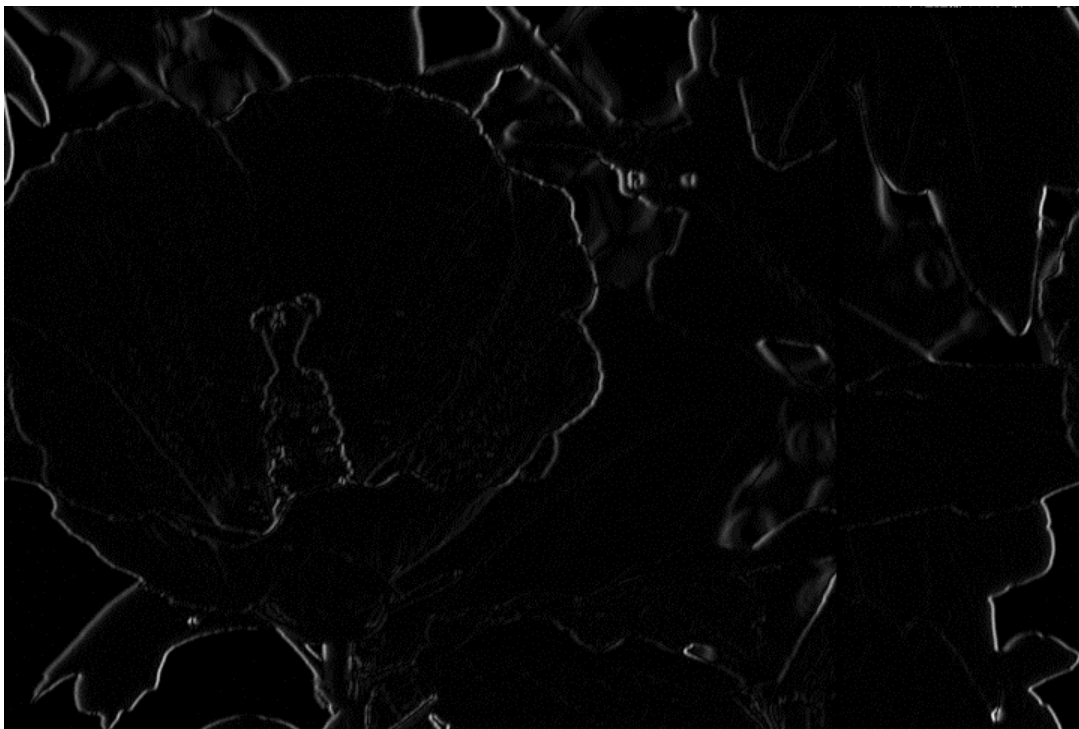
Las Figs. 7-11 ilustran el procedimiento anterior. La Fig. 7 muestra una imagen fuente. La Fig. 8 muestra la imagen de gradiente horizontal  $g_{horizontal}(m,n)$ , obtenida por aplicación de un operador gradiente horizontal a la imagen fuente de la Fig. 7. La Fig. 9 muestra una imagen gradiente vertical  $g_{vertical}(m,n)$ , obtenida por aplicación de un operador gradiente vertical a la imagen fuente de la Fig. 7. La Fig. 10 muestra la imagen de gradiente de magnitud (imagen de bordes) mientras que la Fig. 11 muestra la imagen de bordes binarios (imagen máscara) obtenida por aplicación de umbrales a la imagen de gradiente de magnitud de la Fig. 10.

FIGURA 7  
Imagen fuente (imagen original)



1683-07

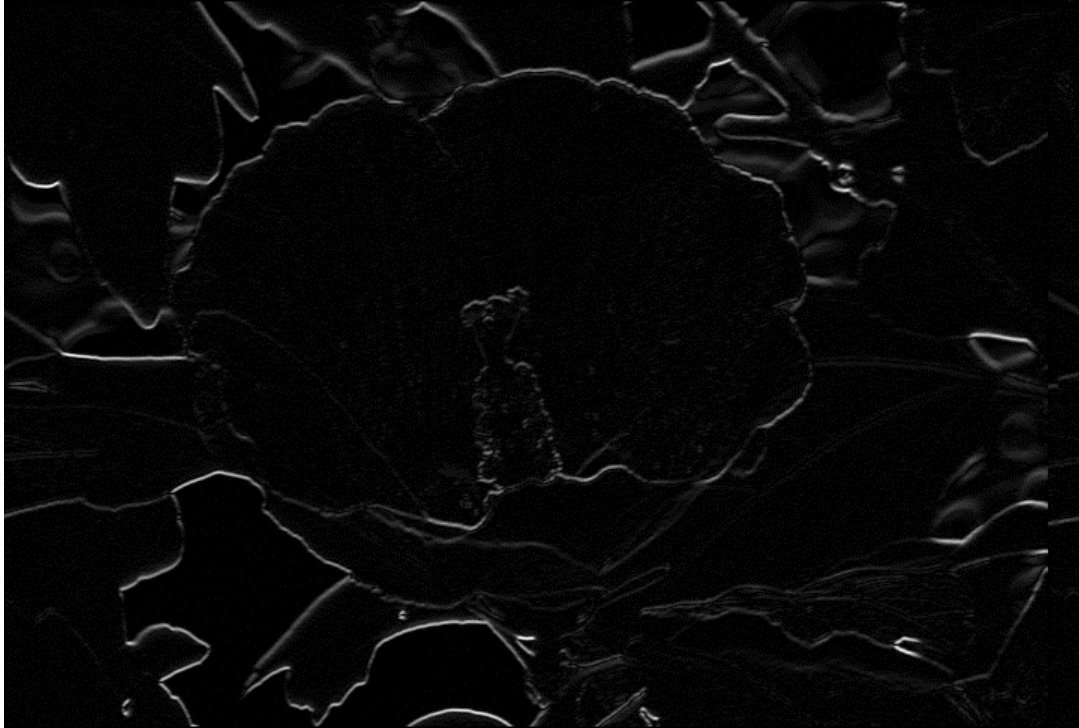
FIGURA 8  
Imagen de gradiente horizontal obtenida por aplicación de un operador  
gradiente horizontal a la imagen fuente de la Fig. 7



1683-08

FIGURA 9

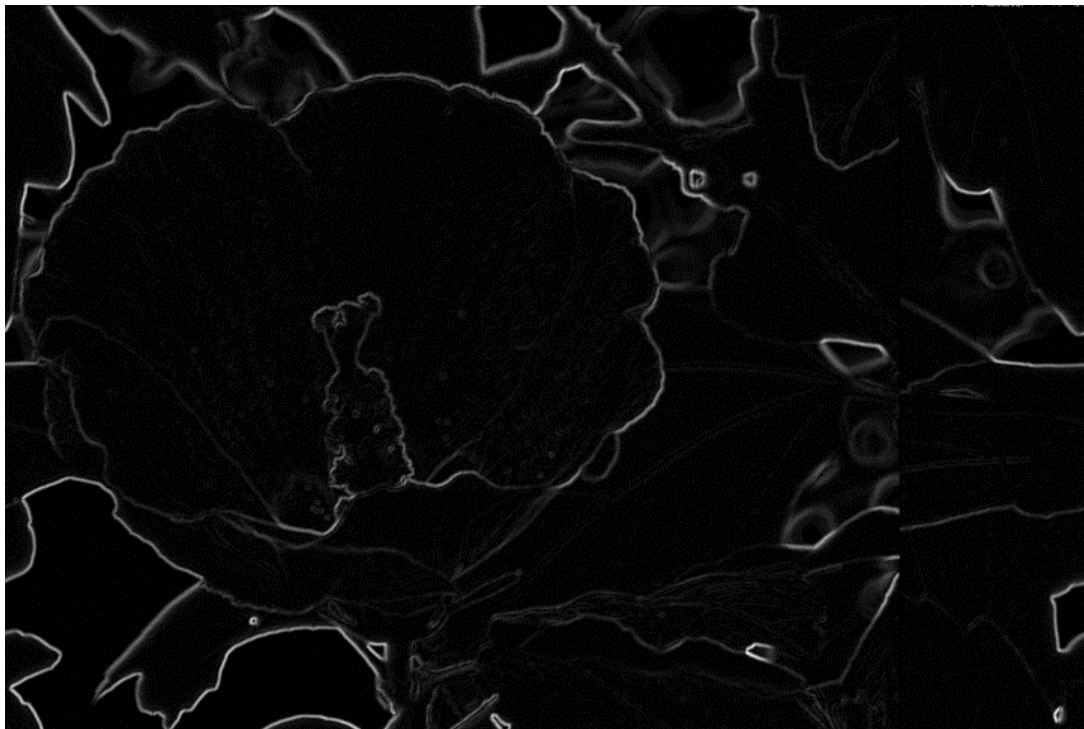
Imagen de gradiente vertical obtenida por aplicación de un operador gradiente vertical a la imagen fuente de la Fig. 7



1683-09

FIGURA 10

Imagen de gradiente de magnitud



1683-10



FIGURA 11

Imagen de bordes binarios (imagen máscara) obtenida por aplicación de umbrales a la imagen de gradiente de magnitud de la Fig. 10



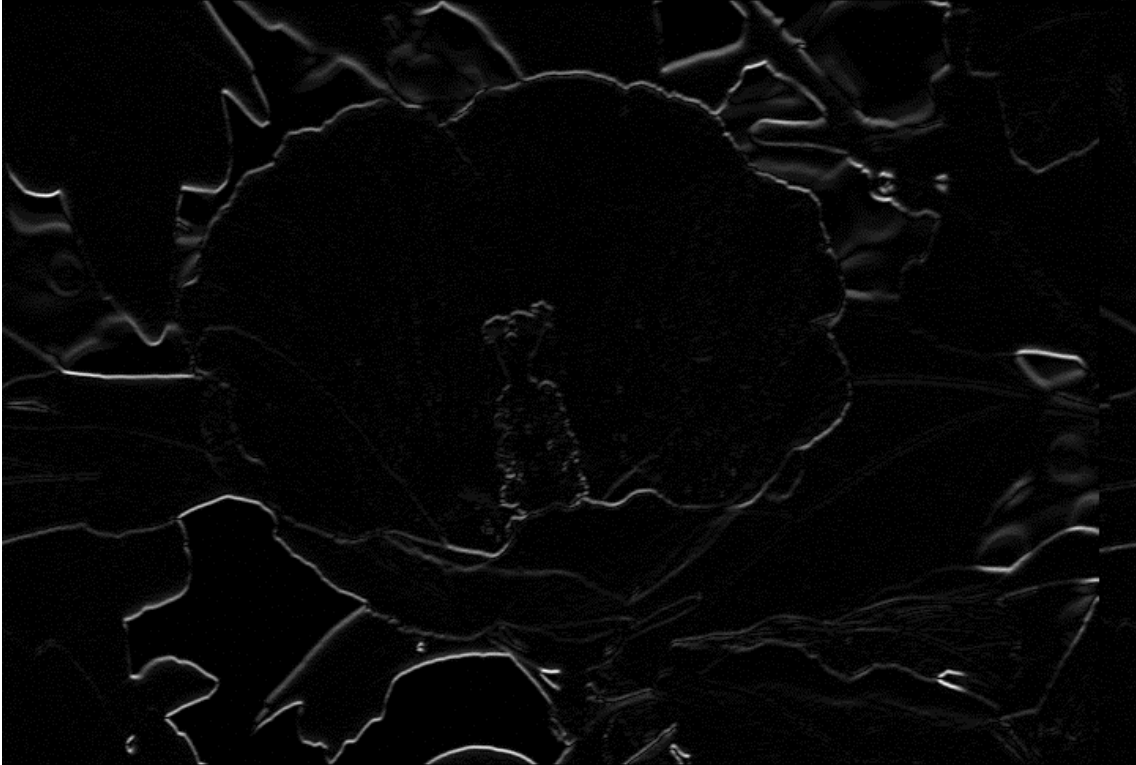
1683-11

Otra solución consiste en utilizar un procedimiento modificado para determinar las zonas de bordes. Por ejemplo, se puede aplicar en primer lugar un operador gradiente vertical a la imagen fuente, produciendo una imagen de gradiente vertical. A continuación se aplica el operador gradiente horizontal a la imagen de gradiente vertical, produciendo una imagen modificada por gradientes sucesivos (imagen de gradiente horizontal y vertical). Finalmente puede aplicarse una operación de fijación de umbrales a la imagen modificada por gradientes sucesivos para determinar las zonas de bordes. Dicho de otro modo, los píxeles de la imagen modificada por gradientes sucesivos que superan un determinado valor umbral, se consideran zonas de bordes. Las Figs. 12-15 ilustran el procedimiento modificado. La Fig. 12 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. 7. La Fig. 13 muestra una imagen modificada por gradientes sucesivos (imagen de gradiente horizontal y vertical) obtenida por aplicación de un operador gradiente horizontal a la imagen de gradiente vertical de la Fig. 12. La Fig. 14 muestra la imagen de bordes binarios (imagen máscara) obtenida por aplicación de umbrales a la imagen modificada por gradientes sucesivos de la Fig. 13.

Hay que observar 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.

FIGURA 12

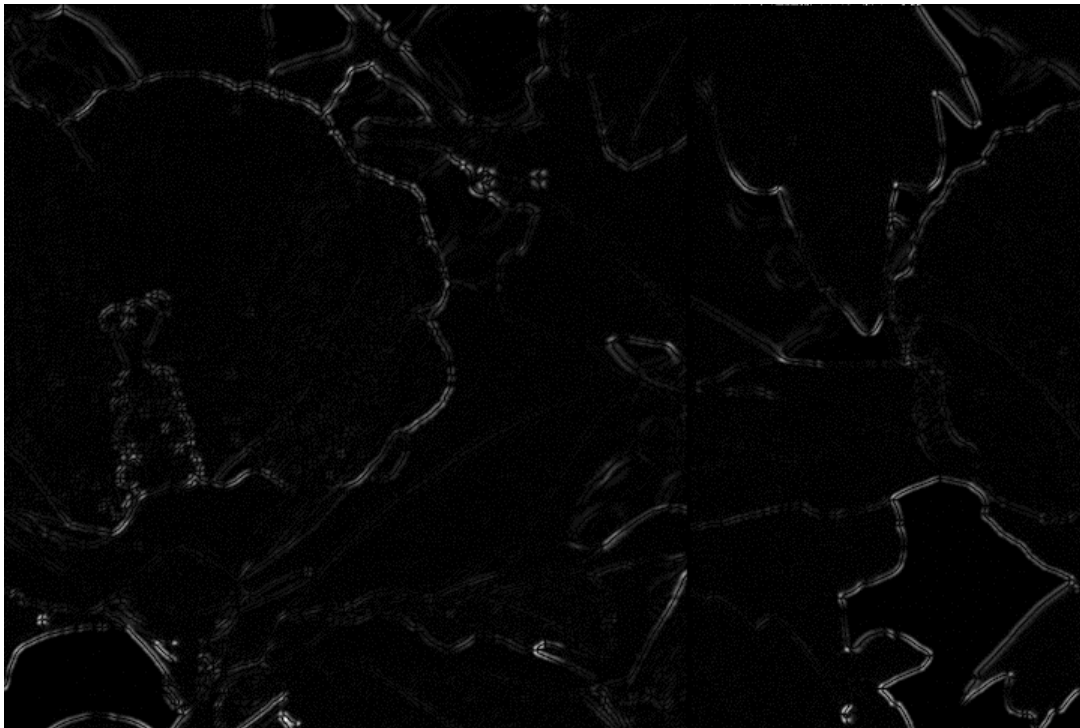
**Imagen de gradiente vertical obtenida por aplicación de un operador gradiente vertical a la imagen fuente de la Fig. 7**



1683-12

FIGURA 13

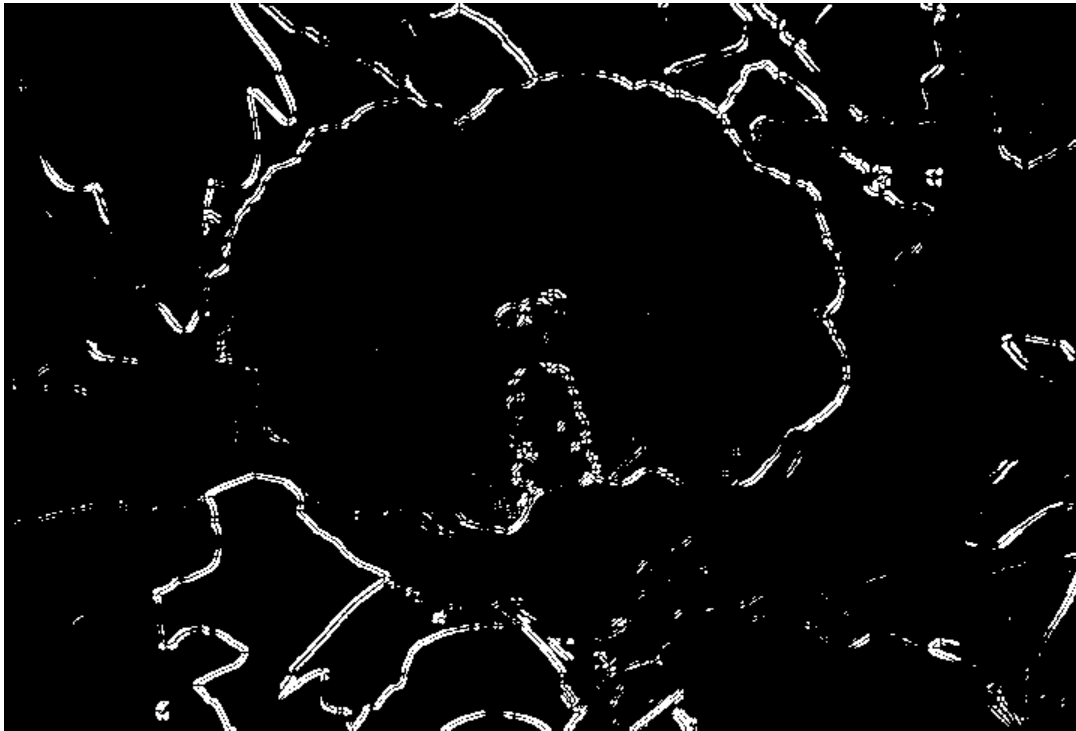
**Imagen modificada por gradientes sucesivos (imagen de gradiente horizontal y vertical) obtenida por aplicación de un operador gradiente horizontal a la imagen de gradiente vertical de la Fig. 12**



1683-13

FIGURA 14

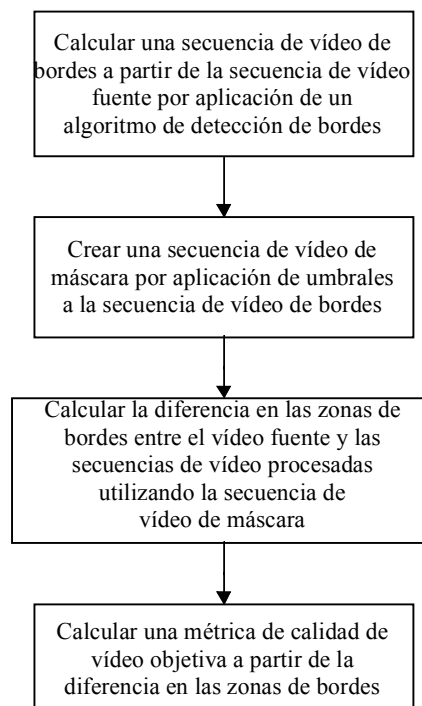
Imagen de bordes binarios (imagen máscara) obtenida por aplicación de umbrales a la imagen modificada por gradientes sucesivos de la Fig. 13



1683-14

FIGURA 15

**Diagrama de bloques de la EPSNR**



1683-15

Así pues, en el modelo, se aplica en primer lugar un operador de detección de bordes, produciendo las imágenes de bordes (Figs. 10 y 13). A continuación se genera una imagen de máscara (imagen de bordes binarios) por aplicación de umbrales a la imagen de bordes (Figs. 11 y 14). 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. 11 y 14 muestran ejemplos de imágenes máscara. Hay que observar que este algoritmo de detección de bordes se aplica a la imagen fuente. Aunque se pueda aplicar el algoritmo de detección de bordes a imágenes procesadas, resulta más preciso aplicarlo a las imágenes fuente. Dado que un vídeo puede contemplarse como secuencia de cuadros o campos, el procedimiento expresado puede aplicarse a cada uno de los cuadros o campos de los vídeos. Como el modelo puede utilizarse para vídeos basados en campos o vídeos basados en cuadros, el término «imagen» se utilizará para indicar ya sea un campo o un cuadro.

A continuación, se calculan las diferencias entre la secuencia de vídeo fuente y la secuencia de vídeo procesada, correspondiente a los píxeles no nulos de la imagen máscara. Dicho de otro modo, se calcula el error cuadrático de las zonas de bordes del cuadro  $l$ -ésimo del siguiente modo:

$$se_e^l = \sum_{i=1}^M \sum_{j=1}^N \{S^l(i, j) - P^l(i, j)\}^2 \quad \text{si } |R^l(i, j)| \neq 0 \quad (75)$$

siendo:

- $S^l(i, j)$ :  $l$ -ésima imagen de la secuencia de vídeo fuente
- $P^l(i, j)$ :  $l$ -ésima imagen de la secuencia de vídeo procesada
- $R^l(i, j)$ :  $l$ -ésima imagen de la secuencia de vídeo máscara
- $M$ : número de filas
- $N$ : número de columnas.

Cuando se implemente este modelo, se puede saltar la generación de la secuencia de vídeo máscara. De hecho, si no se crea la secuencia de vídeo máscara el error cuadrático de las zonas de borde del  $l$ -ésimo cuadro se calcula del siguiente modo:

$$se_e^l = \sum_{i=1}^M \sum_{j=1}^N \{S^l(i, j) - P^l(i, j)\}^2 \quad \text{si } |Q^l(i, j)| \geq t_e \quad (76)$$

siendo:

- $Q^l(i, j)$ :  $l$ -ésima imagen de la secuencia de vídeo de bordes
- $t_e$ : umbral.

Aunque en la ecuación (75) se utilice el error cuadrático medio para calcular la diferencia entre la secuencia de vídeo fuente y la secuencia de vídeo procesada, se puede utilizar cualquier otro tipo de diferencia; por ejemplo, se puede utilizar también la diferencia absoluta. En el modelo presentado a la prueba VQEG Fase II, se otorgó a  $t_e$  el valor 260 y se utilizó el algoritmo modificado de detección de bordes con el operador Sobel.

Este procedimiento se repitió para todas las secuencias de vídeo, calculándose el error cuadrático medio del borde del siguiente modo:

$$mse_e = \frac{1}{K} \sum_{l=1}^L se_e^l \quad (77)$$

siendo:

- $L$ : número de imágenes (cuadros o campos)
- $K$ : número total de píxeles de las zonas de bordes.

Por último, la PSNR de las zonas de bordes (EPSNR) se calculó del siguiente modo:

$$EPSNR = 10 \log_{10} \left( \frac{P^2}{mse_e} \right) \quad (78)$$

siendo:

$P$ : valor de cresta del píxel.

En el modelo, esta EPSNR se utiliza como nota básica de la calidad objetiva de vídeo. La Fig. 15 muestra el diagrama de bloques para el cálculo de la EPSNR.

## 2.2 Ajustes posteriores

### 2.2.1 De énfasis de EPSNR

Cuando el valor de EPSNR es superior a 35, parece que la EPSNR sobreestima la calidad perceptual. Por ello se utiliza el siguiente ajuste lineal de escala por tramos:

$$EPSNR = \begin{cases} EPSNR & \text{si } 0 \leq EPSNR \leq 35 \\ EPSNR \times 0,9 & \text{si } 35 < EPSNR \leq 40 \\ EPSNR \times 0,8 & \text{si } EPSNR > 40 \end{cases} \quad (79)$$

### 2.2.2 Consideración de los bordes borrosos

Se ha observado que cuando los bordes son muy borrosos en los vídeos de baja calidad, los evaluadores tienden a otorgarles notas subjetivas menores. Dicho de otro modo, si las zonas de bordes de la secuencia de vídeo procesada son sustancialmente menores que las de la secuencia de vídeo fuente, los evaluadores otorgan puntuaciones inferiores. Además, se observa que ciertas secuencias de vídeo tienen un número muy pequeño de píxeles con componentes de alta frecuencia. O sea, el número de píxeles de las zonas de bordes es muy pequeño. En consideración a estos problemas, se calculan las zonas de bordes de las secuencias de vídeo fuente y procesada y se modifica la EPSNR del siguiente modo:

$$MEPSNR = \begin{cases} MEPSNR - 60 \times \left( 0,1225 - \left( \frac{EP_{common}}{EP_{src}} \right)^2 \right) \\ MEPSNR \end{cases}$$

si  $EPNSR < 25$  y  $\left( \frac{EP_{common}}{EP_{src}} \right)^2 < 0,35$  (80)

y  $\frac{EP_{hrc}}{EP_{src}} < 0,13$  en los demás casos

siendo:

$MEPSNR$ : EPSNR modificada

$EP_{common}$ : número total de píxeles de bordes comunes en las secuencias de vídeo SRC y HRC (es decir, los píxeles de bordes que aparecen en las mismas posiciones)

$EP_{src}$ : número total de píxeles de bordes de la secuencia de vídeo fuente (SRC).

En algunas secuencias de vídeo,  $EP_{src}$  puede ser muy pequeña. Si  $EP_{src}$  es menor que 10000 píxeles (aproximadamente  $10000/240 = 41,7$  píxeles por cuadro para los vídeos de 525 y 8 s y aproximadamente  $10000/200 = 50$  píxeles por cuadro para los vídeos de 625 y 8 s), el usuario puede reducir el umbral  $t_e$  de la ecuación (76) en 20 hasta que  $EP_{src}$  sea mayor o igual que 10000 píxeles. Si  $EP_{src}$  es menor que 10000 píxeles a pesar de haber reducido  $t_e$  a 80, no se utiliza el ajuste posterior de la ecuación (80). En tal caso, la  $EPSNR$  se calcula con  $t_e = 60$ . Si se toma esta opción, el usuario puede eliminar la condición de que  $EP_{hrc}/EP_{src} < 0,13$  en la ecuación (80).

### 2.2.3 Ajuste

A continuación, las notas objetivas de escala se reajustan de modo que se encuentren entre 0 (indistinguible del vídeo fuente) y 1.

$$VQM = 1 - MEPSNR \times 0,02 \quad (81)$$

Esta VQM se utiliza como nota objetiva del modelo.

## 2.3 Precisión de registro

La precisión de registro recomendada para el modelo es de medio píxel en los vídeos entrelazados, lo que equivale a una precisión de un cuarto de píxel en el formato de vídeo progresivo. Se recomienda encarecidamente la interpolación spline cúbica [Lee y otros, 1998], como mínimo, para calcular los valores de los subpíxeles.

## 2.4 Diagrama de bloques del modelo

La Fig. 16 muestra el diagrama de bloques completo del modelo.

## 3 Datos objetivos

El modelo se aplicó a los datos de vídeo VQEG Fase II<sup>1</sup>. No obstante, tras la presentación del modelo se detectaron errores de registro y del operador. Los datos objetivo presentados en este Anexo son los mismos que los del Informe Final VQEG Fase II. Por consiguiente, si el método descrito en este Anexo se implementa adecuadamente, el usuario puede obtener datos objetivos distintos de los de este Anexo. Los Cuadros 7 y 8 muestran los datos objetivos para conjuntos de datos de vídeo de 525 y 625.

## 4 Conclusión

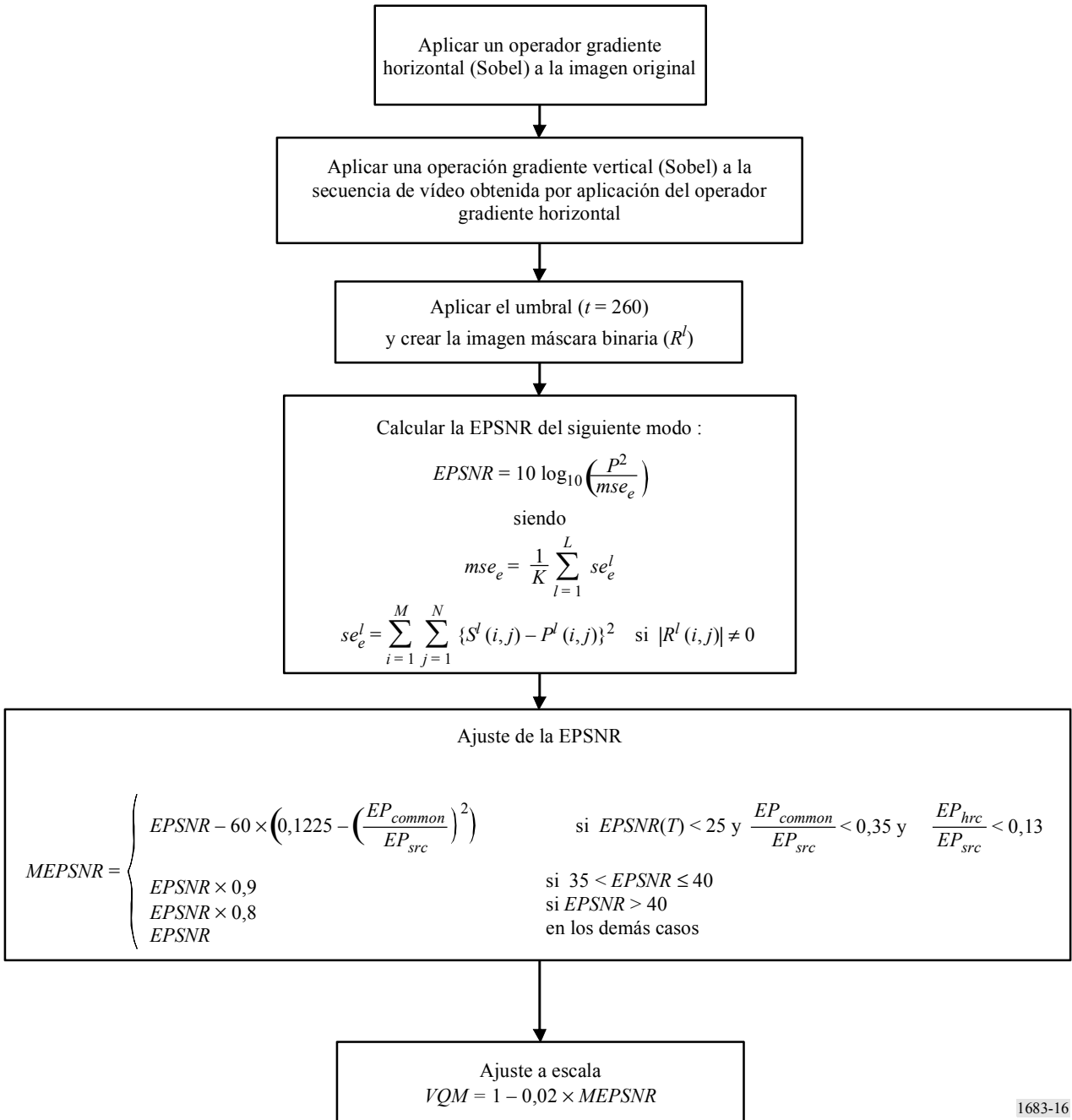
Se propone un nuevo modelo para la medición objetiva de la calidad de vídeo basado en la degradación de los bordes. Este modelo es extremadamente rápido. Una vez generado el mapa de bits, el modelo es varias veces más rápido que la PSNR convencional, lo que supone una mejora importante. Por consiguiente, este modelo es idóneo para aplicaciones que requieran evaluar la calidad de vídeo en tiempo real.

## 5 Referencias Bibliográficas

LEE, C., EDEN, M. y UNSER, M. [1998] High quality image resizing using oblique projection operators. *IEEE Trans. Image Processing*, Vol. 5, 5, p. 679-692.

FIGURA 16

Diagrama de bloques completo del modelo



**CUADRO 7**  
**Matriz VQM 525 (yonsei\_1128c.exe)<sup>(1)</sup>**

SRC (Imagen)	HRC																											
	1		2		3		4		5		6		7		8		9		10		11		12		13		14	
1	1	0,679	4	0,525	7	0,512	10	0,419																				
2	2	0,431	5	0,365	8	0,313	11	0,342																				
3	3	0,558	6	0,452	9	0,340	12	0,305																				
4									13	0,668	17	0,581	21	0,556	25	0,535												
5									14	0,543	18	0,485	22	0,443	26	0,410												
6									15	0,631	19	0,477	23	0,441	27	0,411												
7									16	0,467	20	0,415	24	0,376	28	0,346												
8																	29	0,787	35	0,734	41	0,740	47	0,551	53	0,520	59	0,537
9																	30	0,848	36	0,559	42	0,723	48	0,495	54	0,462	60	0,465
10																	31	0,552	37	0,449	43	0,542	49	0,352	55	0,308	61	0,377
11																	32	0,610	38	0,628	44	0,633	50	0,475	56	0,471	62	0,498
12																	33	0,576	39	0,539	45	0,577	51	0,470	57	0,436	63	0,448
13																	34	0,554	40	0,569	46	0,517	52	0,399	58	0,382	64	0,412

<sup>(1)</sup> Tras la presentación del modelo se detectaron errores de registro y del operador. Los datos objetivo presentados en este Anexo coinciden con los del Informe Final VQEG Fase II. Por consiguiente, si el método descrito en este Anexo se implementa adecuadamente, el usuario puede obtener datos objetivos distintos a los del Cuadro 7.



**CUADRO 8**  
**Matriz VQM 625 (yonsei\_1128c.exe)<sup>(1)</sup>**

SRC (Imagen)	HRC																			
	1		2		3		4		5		6		7		8		9		10	
1			4	0,612	10	0,531	16	0,452			29	0,434			42	0,436			52	0,382
2			5	0,544	11	0,540	17	0,451			30	0,437			43	0,440			53	0,363
3			6	0,572	12	0,571	18	0,497			31	0,479			44	0,478			54	0,418
4			7	0,601	13	0,656	19	0,557			32	0,547			45	0,526			55	0,472
5			8	0,603	14	0,621	20	0,500			33	0,492			46	0,444			56	0,390
6			9	0,591	15	0,520	21	0,483			34	0,469			47	0,461			57	0,423
7							22	0,576			35	0,555					48	0,531	58	0,501
8							23	0,512			36	0,500					49	0,482	59	0,457
9							24	0,507			37	0,487					50	0,468	60	0,436
10							25	0,610			38	0,594					51	0,575	61	0,540
11	1	0,753							26	0,594			39	0,508					62	0,485
12	2	0,643							27	0,556			40	0,550					63	0,496
13	3	0,669							28	0,524			41	0,481					64	0,441

<sup>(1)</sup> Tras la presentación del modelo se detectaron errores de registro y del operador. Los datos objetivo presentados en este Anexo coinciden con los del Informe Final VQEG Fase II. Por consiguiente, si el método descrito en este Anexo se implementa adecuadamente, el usuario puede obtener datos objetivos distintos a los del Cuadro 8.

**Anexo 4****Modelo 3**

## ÍNDICE

	<i>Página</i>
1	Introducción ..... 34
2	Descripción general del sistema IES ..... 35
3	Corrección del desplazamiento y la ganancia ..... 37
3.1	Desplazamiento temporal ..... 37
3.2	Desplazamiento espacial ..... 38
3.3	Ganancia ..... 38
4	Segmentación de la imagen ..... 39
4.1	Regiones planas ..... 39
4.2	Regiones de bordes ..... 39
4.3	Regiones de textura ..... 41
5	Medición objetiva ..... 41
6	Base de datos de los modelos de degradación ..... 41
7	Estimación de los modelos de degradación ..... 42
7.1	Cálculo de $W_i$ ..... 42
7.2	Cálculo de $F_i$ y $G_i$ ..... 43
8	Referencias Bibliográficas ..... 44
	Anexo 4a – Resultados objetivos de las pruebas VQEG-Fase II ..... 45

**1 Introducción**

En este Anexo se presenta una metodología para la evaluación de la calidad de vídeo que utiliza parámetros objetivos basados en la segmentación de imágenes. Las escenas naturales se segmentan en regiones planas, de bordes y de textura, y se asigna un conjunto de parámetros objetivos a cada uno de estos contextos. Se define un modelo perceptual que predice índices subjetivos (Recomendación UIT-R BT.500 – Metodología para la evaluación subjetiva de la calidad de las imágenes de televisión y Recomendación UIT-R BT.802 – Imágenes y secuencias de prueba para las evaluaciones subjetivas de códecs digitales que cursan señales producidas conforme a la Recomendación UIT-R BT.601) a partir del cálculo de la relación entre mediciones objetivas y resultados de pruebas de evaluación subjetiva, aplicadas a un conjunto de escenas naturales procesadas por códecs de vídeo MPEG-2. En este modelo, la relación entre cada parámetro objetivo procesado por varios sistemas de compresión (tales como los códecs MPEG-2 y MPEG-1) y el nivel

de degradación subjetivo, se aproxima por una curva logística, obteniéndose como resultado un nivel de degradación estimado para cada parámetro. El resultado final se obtiene a partir de una combinación lineal de los niveles de degradación estimados, siendo el peso de cada nivel de degradación proporcional a su fiabilidad estadística.

En el § 2, se presenta una descripción general del sistema de evaluación de imágenes por segmentación (CPqD-IES, *image evaluation based on segmentation*). En el § 3, se describen los pasos para registrar las desalineaciones espaciales y temporales, así como la corrección de la ganancia. En el § 4, se presenta el algoritmo de segmentación de imágenes en regiones planas, de bordes y de textura. En el § 5, se describen las mediciones objetivas realizadas en cada una de las regiones y sobre cada una de las componentes de imagen. En el § 6 se describe cómo se construyó la base de datos de los modelos de degradación. En este apartado se describe asimismo el cálculo de los parámetros. En el § 7 se describe la estimación del índice de calidad de vídeo a partir de los parámetros de las bases de datos de los modelos de degradación. En el Anexo 4a se presentan los resultados de los valores del índice de calidad de vídeo (VQR, *video quality rating*) estimados durante las pruebas VQEG Fase II<sup>1</sup>.

## 2 Descripción general del sistema IES

La Fig. 17 muestra un esquema general del algoritmo CPqD-IES para una escena natural. Una escena natural se representa por una escena original  $\mathbf{O}$  (de referencia) y una escena degradada  $\mathbf{I}$ , resultante de una operación códec aplicada a  $\mathbf{O}$ . Se aplican a  $\mathbf{I}$  correcciones de ganancia y de desplazamiento para crear una escena degradada corregida  $\mathbf{I}'$ , tal que cada cuadro  $f$  de  $\mathbf{I}'$  corresponda a un cuadro de referencia  $f$  de  $\mathbf{O}$ , para  $f = 1, 2, \dots, n$  (véase el § 3.2).

Las escenas de entrada  $\mathbf{I}$  y  $\mathbf{O}$  al algoritmo CPqD-IES tienen el formato YCbCr4:2:2 con arreglo a la Recomendación UIT-R BT.601 – Parámetros de codificación de televisión digital para estudios con formatos de imagen normal 4:3 y de pantalla ancha 16:9.

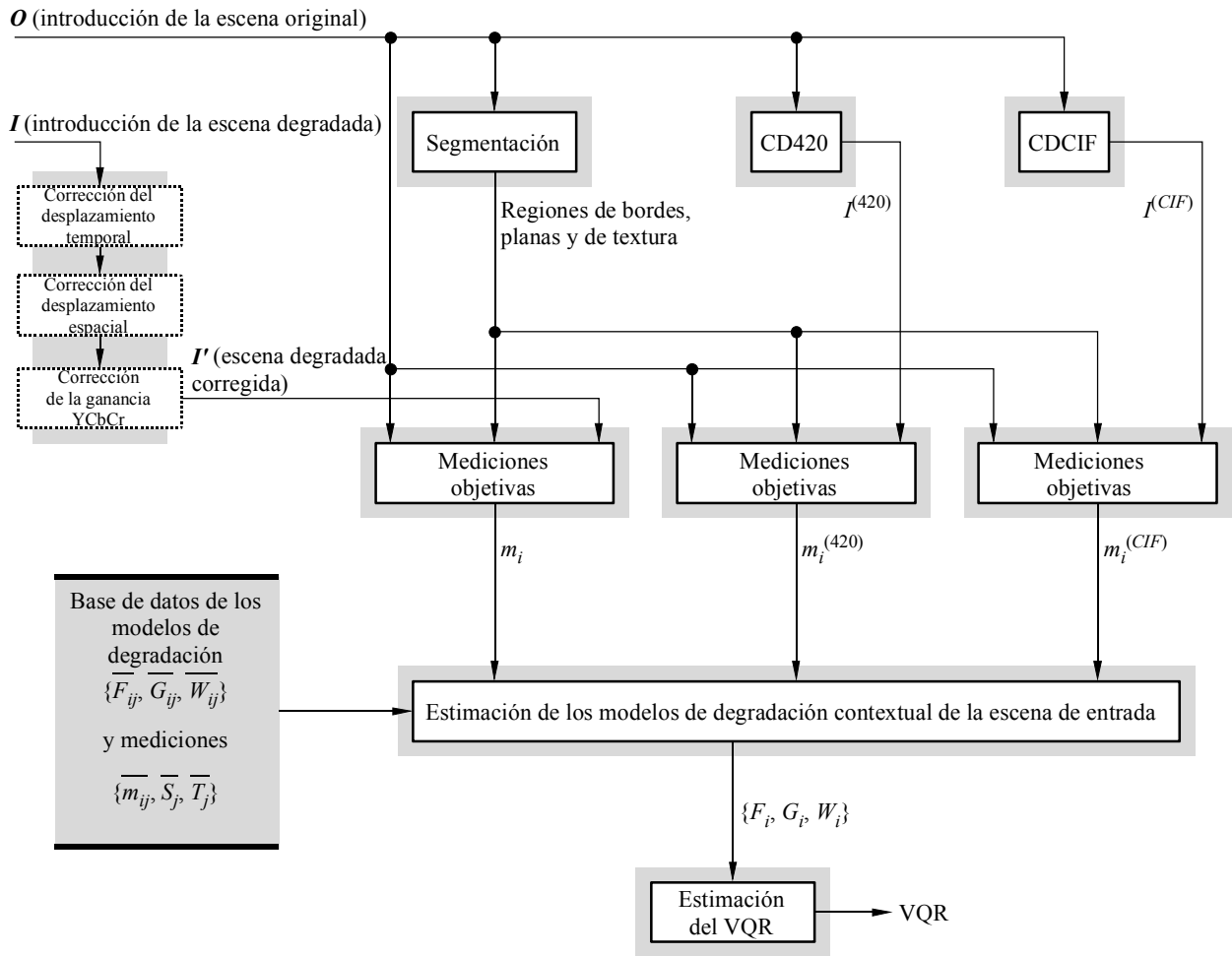
La componente Y de cada cuadro de imagen  $f$  de  $\mathbf{O}$  se segmenta en tres categorías: regiones de textura, regiones de borde y regiones planas (véase el § 4). Una medición objetiva se calcula en base a la diferencia entre los cuadros correspondientes de  $\mathbf{O}$  e  $\mathbf{I}'$ , para cada uno de estos contextos y para cada componente de imagen YCbCr, formando un conjunto de 9 mediciones objetivas  $\{m_1, m_2, \dots, m_9\}$  para cada cuadro de imagen  $f$  (véase el § 5). Cada medición objetiva  $m_i$ ,  $i = 1, 2, \dots, 9$ , produce un nivel de degradación contextual  $L_i$  basado en su modelo de estimación de degradación, representado por:

$$L_i = 100 / \left[ 1 + \left( \frac{F_i}{m_i} \right)^{G_i} \right] \quad (82)$$

siendo  $F_i$  y  $G_i$  dos parámetros calculados (véase el § 7) a partir de la base de datos de los modelos de degradación (véase el § 6), de los atributos espacial  $S$  y temporal  $F$  (véase el § 5), y de las mediciones objetivas  $m_i^{(420)}$  y  $m_i^{(CIF)}$  del cuadro  $f$ , resultantes de las operaciones códec CD420 y CDCIF aplicadas a  $\mathbf{O}$  (véase el § 7). Las dos referencias de los códecs de los modelos de degradación CD420 (codificador/decodificador MPEG-2 4:2:0) y CDCIF (codificador/decodificador MPEG-1 CIF), se basan completamente en rutinas sacadas directamente de MPEG2 (Recomendación UIT-T H.262 – Tecnología de la información – Codificación genérica de imágenes en movimiento e información de audio asociada: Vídeo) y MPEG1 [ISO/CEI, 1992], disponibles en [www.mpeg.org/MPEG/MSSG](http://www.mpeg.org/MPEG/MSSG). En la implementación actual del algoritmo CPqD-IES, estas rutinas funcionan en intramodo utilizando un paso de cuantificación de valor fijo 16. Es importante observar que CD420 y CDCIF no introducen desplazamiento ni diferencia de ganancia con respecto a  $\mathbf{O}$ .

FIGURA 17

## Esquema general del algoritmo CPqD-IES



CIF: Formato intermedio común

1683-17

El índice de calidad de vídeo del cuadro  $f$ ,  $VQR_f$ , se obtiene por combinación lineal de los niveles de degradación contextuales  $L_i$ ,  $i = 1, 2, \dots, 9$ , del siguiente modo:

$$VQR_f = \sum_{i=1}^9 W_i \cdot L_i \quad (83)$$

siendo  $W_i$  el peso del nivel de degradación  $L_i$  correspondiente a esta escena natural en particular, el cual se calcula como se describe en el § 7.

A continuación, la secuencia de valores  $VQR_1, VQR_2, \dots, VQR_n$  se transforma mediante un filtro de mediana de tamaño 3 en otra secuencia  $VQR'_1, VQR'_2, \dots, VQR'_n$ , excluyendo el cálculo del valor de la mediana en la vecindad-1 de  $VQR_1$  y  $VQR_n$ . Durante el filtrado de mediana, el algoritmo evita la repetición de dos valores de mediana consecutivos. O sea, si el valor de la mediana  $VQR'_{f-1}$  calculado en la vecindad-1 de  $VQR_f$  es igual al valor de mediana  $VQR'_{f-2}$  calculado en la vecindad-1 de  $VQR_{f-1}$ , el algoritmo selecciona  $VQR'_{f-1}$  como valor mínimo calculado en la vecindad-1 de  $VQR_f$ .

Este algoritmo puede describirse del siguiente modo.

- 1) Para cada  $f$  desde 2 a  $n - 1$
- 2) Calcular  $med$ , valor de la mediana entre  $VQR_{f-1}$ ,  $VQR_f$ ,  $VQR_{f+1}$
- 3) Si  $med = VQR'_{f-2}$  entonces
- 4) Calcular  $VQR'_{f-1}$  como valor mínimo entre  $VQR_{f-1}$ ,  $VQR_f$ ,  $VQR_{f+1}$
- 5) De lo contrario
- 6)  $VQR'_{f-1} \leftarrow med$ .

Así pues, el  $VQR$  final es el promedio de los valores  $VQR'_f$

$$VQR = \frac{1}{n-2} \cdot \sum_{f=1}^{n-2} VQR'_f \quad (84)$$

Las ecuaciones (82), (83) y el algoritmo anterior describen el proceso de estimación del  $VQR$  a partir de los modelos de degradación contextual  $\{F_i, G_i, W_i\}$  y de las mediciones objetivas  $m_i, i = 1, 2, \dots, 9$ . En los apartados siguientes se completa la descripción de este método con los detalles correspondientes a los demás bloques de la Fig. 17.

### 3 Corrección del desplazamiento y la ganancia

#### 3.1 Desplazamiento temporal

El desplazamiento temporal,  $dt$ , es un entero comprendido entre  $-2$  y  $2$ . Las escenas de entrada con desplazamientos temporales fuera de este rango no se consideran. Sea  $I_{dt}$  la escena  $I$  degradada con un desplazamiento de  $f$  cuadros. Se calcula un coeficiente de desemejanza entre  $O$  y cada escena desplazada  $I_{dt}$ . El desplazamiento con menor coeficiente de desemejanza se utiliza como desplazamiento temporal y, en el siguiente cálculo, la salida  $I_{dt}$  se desplaza con respecto a  $I$  en esta magnitud. El coeficiente de desemejanza entre  $O$  e  $I_{dt}$  se obtiene como se indica a continuación, siendo  $n$  el número de cuadros de la intersección temporal entre ellos:

- 1)  $\xi_T \leftarrow 0$
- 2) Para cada  $f$  de 1 a  $n$
- 3) Calcular  $S_b$
- 4) Calcular  $S'_b$
- 5) Calcular  $D_b$
- 6) Calcular  $\mu$ , valor medio de los píxeles de  $D_b$
- 7)  $\xi_T \leftarrow \xi_T + (\mu/n)$
- 8) Devolver  $\xi_T$  (coeficientes de desemejanza entre  $O$  e  $I_{dt}$ ).

Siendo:

$S_b$ : magnitud del gradiente de Sobel de la componente  $Y$  del  $f$ -ésimo cuadro de  $O$

$S'_b$ : magnitud del gradiente de Sobel de la componente  $Y$  del  $f$ -ésimo cuadro de  $I_{dt}$

$D_b$ : diferencia absoluta en píxeles entre  $S_b$  y  $S'_b$ .

### 3.2 Desplazamiento espacial

El desplazamiento espacial ( $d_x, d_y$ ) será uno de los siguientes desplazamientos enteros verticales u horizontales,  $d_x = -6, -5, \dots, 6$  y  $d_y = -6, -5, \dots, 6$ . Sea  $I_{dx,dy}$  la escena degradada  $I_{dt}$  con todos sus cuadros desplazados de ( $d_x, d_y$ ) píxeles. Se calcula un coeficiente de desemejanza entre  $O$  e  $I_{dx,dy}$ . El desplazamiento espacial con menor desemejanza se utiliza como desplazamiento espacial y, en el siguiente cálculo, la salida  $I_{dx,dy}$  se desplaza con respecto a  $I_{dt}$  en esta magnitud.

La desemejanza entre  $O$  e  $I_{dx,dy}$  se describe a continuación:

- 1)  $\xi_S \leftarrow 0 ; c \leftarrow 0$
- 2) Para cada  $f$  entre 1 y  $n$
- 3) Para  $x$  entre  $x_0$  y  $(x_0 + w/4)$
- 4) Para  $y$  entre  $y_0$  y  $(y_0 + h/4)$
- 5)  $\xi_S \leftarrow \xi_S + |Y(4x,4y) - Y'(4x + dx, 4y + dy)| +$   
 $+ |Cb(4x,4y) - Cb'(4x + dx, 4y + dy)| +$   
 $+ |Cr(4x,4y) - Cr'(4x + dx, 4y + dy)|$
- 6)  $c \leftarrow c + 3$
- 7)  $\xi_S \leftarrow \xi_S / c$
- 8) Devolver  $\xi_S$  (coeficiente de desemejanza entre  $O$  e  $I_{dx,dy}$ ).

Siendo:

$w \times h$ : tamaño de la zona de intersección entre  $O$  e  $I_{dx,dy}$

$Y(x, y), Cb(x, y), Cr(x, y)$ : valores de las componentes de imagen del cuadro  $f$  de  $O$  para el píxel  $(x, y)$ ;

$Y'(x + dx, y + dy)$

$Cb'(x + dx, y + dy)$

$Cr'(x + dx, y + dy)$ : valores de las componentes de imagen del cuadro  $f$  de  $I_{dx,dy}$  para el píxel  $(x + dx, y + dy)$ .

### 3.3 Ganancia

La ganancia de amplitud entre  $O$  e  $I_{dx,dy}$  se calcula para cada componente de imagen  $Y, C_B$  y  $C_R$ , por separado. El algoritmo calcula el promedio de las ganancias para los  $n$  cuadros e introduce las oportunas correcciones para cada componente de imagen. La salida  $I'$  es la escena degradada utilizada en todos los cálculos subsiguientes. La ganancia de amplitud entre una componente de imagen  $C$  del cuadro  $f$  de  $I_{dx,dy}$  con respecto a la misma componente  $C$  del cuadro  $f$  de  $O$  se obtiene difuminando ambas imágenes  $C'$  y  $C$ , por medio de un filtro Gaussiano [González y Woods, 1992] cuyo núcleo sea:

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}$$

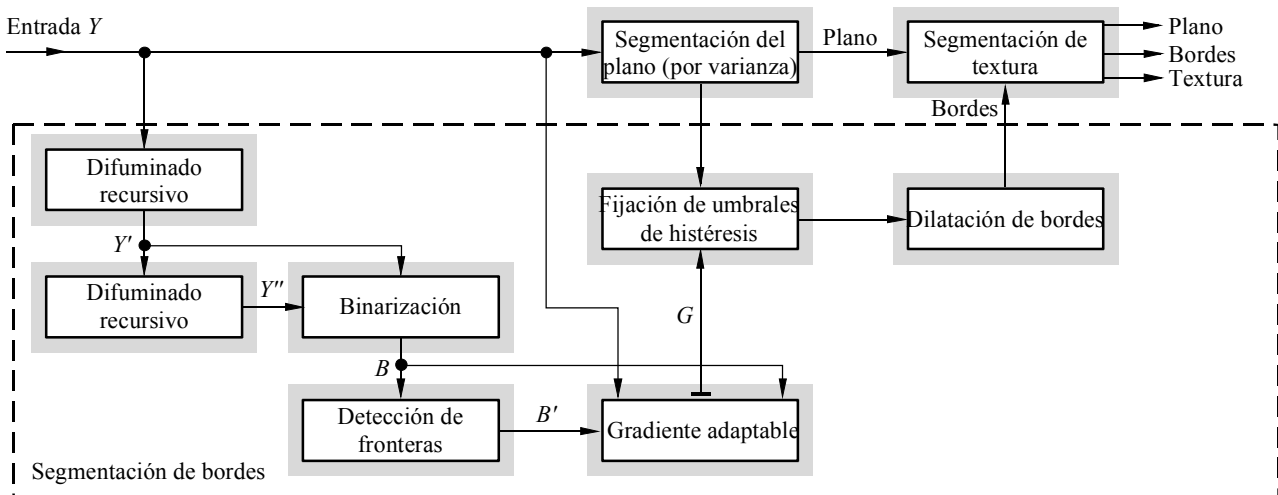
y calculando la relación entre la suma de sus valores de píxel en las imágenes difuminadas. Sólo se considera 1 de cada 16 píxeles (barriendo las imágenes de componentes difuminadas con incrementos verticales y horizontales de 4 píxeles, como en el algoritmo de cálculo de  $\xi_S$  descrito en el § 3.2).

#### 4 Segmentación de la imagen

Inicialmente, el algoritmo de segmentación clasifica cada píxel de la componente  $Y$  de un cuadro determinado  $f$  de la escena original  $O$  en regiones planas y no planas. Además, el algoritmo aplica a  $Y$  un detector de bordes, quedando definida la región de bordes por aquellos que quedan dentro de los límites de la región plana. La región de textura está compuesta por los restantes píxeles de la imagen  $Y$  (véase la Fig. 18).

FIGURA 18

Diagrama de bloques del proceso de segmentación



1683-18

La segmentación se computa para cada cuadro de la componente  $Y$  de la escena original de entrada  $O$ . Para las componentes  $C_B$  y  $C_R$ , las regiones se definen por la posición de los píxeles en la componente  $Y$ , tras la generación de muestras interpoladas (up-sampling) en  $C_B$  y  $C_R$ .

##### 4.1 Regiones planas

La varianza del brillo de cada píxel de  $Y$  se calcula en la vecindad de  $5 \times 5$  píxeles a su alrededor. Se fijan umbrales para la varianza de la imagen tales que los píxeles cuyo valor de varianza sea inferior a  $25^2$  se clasifiquen como pertenecientes a la región plana. Este proceso crea pequeños componentes de píxeles que no corresponden a la región de textura. Se aplica un filtro de mediana de  $3 \times 3$  para suprimir estas pequeñas componentes. Por último, la imagen binaria de las regiones planas se somete a una dilatación morfológica utilizando un elemento de estructuración circular de 11 píxeles de diámetro [González y Woods, 1992].

##### 4.2 Regiones de bordes

Se aplica un filtrado recursivo a  $Y$ , creando una primera imagen difuminada  $Y'$  que se aplica a continuación a  $Y'$  creando una segunda imagen difuminada  $Y''$ . Cada filtrado recursivo consta de cuatro rásteres de la señal de entrada. Este algoritmo se describe a continuación para la componente de imagen  $Y$  de un cuadro de la escena de entrada  $O$ .

- 1) Para  $y$  comprendida entre 0 y  $(h - 1)$
- 2) Para  $x$  comprendida entre 0 y  $(w - 2)$
- 3) 
$$Y(x + 1, y) \leftarrow Y(x, y) + 0,7[Y(x + 1, y) - Y(x, y)]$$

- 4) Para  $y$  comprendida entre 0 y  $(h - 1)$
- 5) Para  $x$  comprendida entre  $(w - 1)$  y 1
- 6)  $Y(x - 1, y) \leftarrow Y(x, y) + 0,7[Y(x - 1, y) - Y(x, y)]$
- 7) Para  $x$  comprendida entre 0 y  $(w - 1)$
- 8) Para  $y$  comprendida entre 0 y  $(h - 2)$
- 9)  $Y(x, y + 1) \leftarrow Y(x, y) + 0,7[Y(x, y + 1) - Y(x, y)]$
- 10) Para  $x$  comprendida entre 0 y  $(w - 1)$
- 11) Para  $y$  comprendida entre  $(h - 1)$  y 1
- 12)  $Y(x, y + 1) \leftarrow Y(x, y) + 0,7[Y(-1) - Y(x, y)]$
- 13) Guardar la imagen  $Y$  en la imagen  $Y'$ .

Siendo:

$Y(x, y)$ : brillo del píxel  $(x, y)$   
 $h$ : número de líneas de  $Y$   
 $w$ : número de columnas de  $Y$ .

En la segunda aplicación del algoritmo anterior se crea  $Y''$ . Se crea una imagen binaria  $B$  a partir de  $Y'$  e  $Y''$ :

$$B(x, y) = \begin{cases} 1 & \text{si } Y'(x, y) \geq Y''(x, y) \\ 0 & \text{en los demás casos} \end{cases} \quad (85)$$

A continuación, el algoritmo identifica los píxeles fronterizos de las regiones de  $B$  cuyo valor de píxel sea 1 mediante la creación de una segunda imagen binaria  $B'$ :

$$B'(x, y) = \begin{cases} 1 & \text{si } B(x, y) = 1 \text{ y } B(x', y') = 0 \text{ para cualquier píxel } (x', y') \in N_8(x, y) \\ 0 & \text{en los demás casos} \end{cases} \quad (86)$$

siendo  $N_8(x, y)$  el conjunto de píxeles  $(x', y')$  de la vecindad de  $3 \times 3$  de  $(x, y)$  (o sea, sus 8 vecinos).

Se aplica un filtrado de gradiente adaptable a  $Y$  restringido a los píxeles en los que  $B'(x, y) = 1$ :

$$G(x, y) = \begin{cases} |\mu_1 - \mu_0| & \text{si } B'(x, y) = 1 \\ 0, & \text{en los demás casos} \end{cases} \quad (87)$$

siendo:

$\mu_1$ : valor medio de  $Y(x', y')$ , para todo  $(x', y') \in N_8(x, y)$  de modo que  $B(x', y') = 1$

$\mu_0$ : valor medio de  $Y(x', y')$ , para todo  $(x', y') \in N_8(x, y)$  de modo que  $B(x', y') = 0$ .

Obsérvese que este algoritmo utiliza  $B$  en vez de  $B'$  para calcular los valores medios  $\mu_1$  y  $\mu_0$ .

Se aplica un umbral de histéresis [Trucco y Verri, 1998] a  $G$  restringido a los píxeles clasificados en el § 4.1 como pertenecientes a la región plana. El umbral inferior es 30 y el superior 40. El algoritmo identifica primero los píxeles de  $G$ , tales que  $G(x, y) > 40$ , y aplica a continuación un algoritmo de crecimiento de región a lo largo de las líneas de  $G$  utilizando estos píxeles como semillas y restringiendo el crecimiento a los píxeles de la misma línea cuya  $G(x, y) > 30$ . Todos los componentes de 4 píxeles conectados con menos de 6 píxeles se eliminan de este resultado. La imagen binaria final se dilata mediante un elemento de estructuración circular con un diámetro de 5 píxeles ignorando la restricción para la región plana. En esta dilatación, los píxeles con valor 1 se clasifican como pertenecientes a la región de bordes.



### 4.3 Regiones de textura

La región de textura consiste en los píxeles de  $Y$  que no fueron clasificados como pertenecientes a la región de bordes ni a la región plana de los apartados anteriores.

## 5 Medición objetiva

Sea  $S_b$  la imagen de la magnitud del gradiente de Sobel calculado para una componente determinada ( $Y$ ,  $C_B$  o  $C_R$ ) de un cuadro determinado  $f$  de la escena original  $\mathbf{O}$ , y  $S'_b$ , la imagen de la magnitud del gradiente de Sobel de la misma componente del cuadro  $f$  de la escena degradada  $\mathbf{I}'$ . Se calcula la diferencia absoluta en píxeles  $D_b$  de la imagen entre  $S_b$  y  $S'_b$  y se considera la región  $\mathfrak{R}$  de píxeles de  $D_b$  que pertenece a un contexto determinado (plano, borde o textura). Se define la diferencia absoluta de Sobel (ASD, *absolute Sobel difference*) para esta componente de imagen y contexto como el promedio de los valores de píxel de  $D_b$  restringidos a  $\mathfrak{R}$ .

De este procedimiento se obtiene un conjunto de nueve mediciones objetivas  $\{m_1, m_2, \dots, m_9\}$  para cada cuadro de imagen  $f$ ,  $f = 1, 2, \dots, n$ , considerando los tres contextos y las tres componentes de imagen.

Se aplica el mismo proceso para crear mediciones objetivas  $\{m_1^{(420)}, m_2^{(420)}, \dots, m_9^{(420)}\}$  y  $\{m_1^{(CIF)}, m_2^{(CIF)}, \dots, m_9^{(CIF)}\}$  para el cuadro  $f$  con respecto a las operaciones MPEG-2 4:2:0 y MPEG-1 CIF CODEC sobre  $\mathbf{O}$  (véase la Fig. 17). Estas mediciones se utilizan como referencia junto con los atributos espacial  $S$  y temporal  $T$  para determinar el modelo de degradación contextual para  $\mathbf{I}'$  (véase el § 7). El atributo temporal  $T$  es el valor medio de la diferencia absoluta en píxeles entre las segmentaciones de los cuadros  $f$  y  $f-1$ , normalizada a  $[0,1]$ . El atributo espacial  $S$  se define como la relación  $m_7^{(CIF)}/m_7^{(420)}$ , normalizada a  $[0,1]$ , siendo  $m_7^{(CIF)}$  y  $m_7^{(420)}$  las correspondientes ASD para la región de textura de la componente  $Y$  del cuadro  $f$ .

## 6 Base de datos de los modelos de degradación

El sistema IES utiliza una base de datos de modelos de degradación para escenas distintas a la escena de referencia  $\mathbf{O}$  a fin de estimar el índice de calidad de vídeo de  $\mathbf{I}'$ . Esta base de datos está integrada por información relativa a 12 escenas de 60 Hz que representan diversos grados de movimiento (escenas dinámicas y estáticas), naturaleza (escenas reales y sintéticas), y contexto (cantidad de píxeles de textura, plano y bordes). Esta base de datos se creó de la siguiente manera.

Se calcularon los valores medios de las mediciones objetivas  $\{\bar{m}_1, j, \bar{m}_2, j, \dots, \bar{m}_9, j\}$ ,  $\{\bar{m}_1^{(420)}, \bar{m}_2^{(420)}, \dots, \bar{m}_9^{(420)}\}$  y  $\{\bar{m}_1^{(CIF)}, \bar{m}_2^{(CIF)}, \dots, \bar{m}_9^{(CIF)}\}$  para los cuadros de cada escena  $j$ ,  $j = 1, 2, \dots, 12$ . Los valores de  $S_j$  y  $T_j$  se calcularon como promedio de los atributos espacial y temporal, calculados como se describe en el § 5, sobre los cuadros de cada escena  $j$ . Todas las escenas degradadas de la base de datos se sometieron además a evaluación subjetiva, obteniéndose para cada escena  $j$  un nivel de degradación subjetivo,  $SL_j$ , normalizado entre 0% y 100%.

De acuerdo con la ecuación (82), cada medición objetiva  $m_{i,j}$ ,  $i = 1, 2, \dots, 9$  y  $j = 1, 2, \dots, 12$ , se relaciona con un nivel de degradación contextual  $L_{i,j}$ . Se determinaron los valores  $F_{i,j}$  y  $G_{i,j}$  de la ecuación (82) para cada escena  $j$  minimizando el MSE esperado  $E[(\overline{SL}_j - \bar{L}_{i,j})^2]$ . Además, se calcularon los valores de  $W_{i,j}$  de la ecuación (83) para minimizar el MSE esperado:

$$E \left[ \left( \overline{SL}_j - \sum_{i=1}^9 \bar{W}_{i,j} \cdot \bar{L}_{i,j} \right)^2 \right] \quad (88)$$

Por último, se definió la base de datos de los modelos de degradación con nueve conjuntos de parámetros  $\{\bar{F}_{i,j}, \bar{G}_{i,j}, \bar{W}_{i,j}, \bar{S}_j, \bar{T}_j\}$ ,  $i = 1, 2, \dots, 9$  para cada escena  $j$ ,  $j = 1, 2, \dots, 12$ . El Cuadro 9 contiene los valores de  $\bar{S}_j$  y  $\bar{T}_j$  para calcular los atributos  $\{\bar{F}_{i,j}, \bar{G}_{i,j}, \bar{W}_{i,j}\}$ .

CUADRO 9

Atributos temporales  $T$  y espaciales  $S$ 

Escena $j$	$T$ (temporal)	$S_Y$ (espacial $Y$ )	$S_{Cb}$ (espacial $C_B$ )	$S_{Cr}$ (espacial $C_R$ )
1	27,01	36,79	25,20	38,01
2	25,33	26,08	5,93	67,99
3	45,54	60,97	10,28	28,75
4	36,40	30,47	6,46	63,07
5	32,02	72,50	11,72	15,78
6	12,63	84,22	2,85	12,94
7	28,38	61,53	11,08	27,39
8	10,19	46,08	5,45	48,47
9	0,01	5,89	5,07	89,03
10	7,26	4,75	2,00	93,25
11	7,60	69,16	9,41	21,43
12	14,27	69,61	3,89	26,50

## 7 Estimación de los modelos de degradación

Los modelos de degradación contextual para un determinado cuadro  $f$  de  $I'$  constan de los parámetros  $\{F_i, G_i, W_i\}$  de las ecuaciones (82) y (83),  $i = 1, 2, \dots, 9$ . En este apartado se describe el cálculo de dichos parámetros utilizando las escenas degradadas  $I^{(420)}$  e  $I^{(CIF)}$  como referencia.

### 7.1 Cálculo de $W_i$

Las distancias locales contextuales  $D_{i,j}$  entre un cuadro  $f$  de las escenas degradadas,  $I^{(420)}$  e  $I^{(CIF)}$ , y cada escena  $j$  de la base de datos se definen del siguiente modo:

$$\begin{aligned}
 \bar{L}_{i,j}^{(420)} &= 100 / \left[ 1 + \left( \bar{F}_{i,j} / \bar{m}_i^{(420)} \right) \bar{G}_{i,j} \right] \\
 \bar{L}_{i,j}^{(CIF)} &= 100 / \left[ 1 + \left( \bar{F}_{i,j} / \bar{m}_i^{(CIF)} \right) \bar{G}_{i,j} \right] \\
 L_{i,j}^{(420)} &= 100 / \left[ 1 + \left( \bar{F}_{i,j} / m_i^{(420)} \right) \bar{G}_{i,j} \right] \\
 L_{i,j}^{(CIF)} &= 100 / \left[ 1 + \left( \bar{F}_{i,j} / m_i^{(CIF)} \right) \bar{G}_{i,j} \right]
 \end{aligned} \tag{89}$$

$L_{i,j}^{(420)}$  y  $L_{i,j}^{(CIF)}$  son los niveles de degradación estimados de la escena de entrada  $\mathbf{O}$ , calculados con los parámetros  $\bar{F}_{i,j}$  y  $\bar{G}_{i,j}$ , en el contexto  $i$ , de las escenas  $j$  de la base de datos.

$$D_{i,j} = \frac{1}{2} \cdot \left( \left| L_{i,j}^{(420)} - \bar{L}_{i,j}^{(420)} \right| + \left| L_{i,j}^{(CIF)} - \bar{L}_{i,j}^{(420)} \right| \right) \quad (90)$$

Este algoritmo permite encontrar el conjunto  $\Omega$  de las seis escenas más próximas de la base de datos a partir de la distancia  $D_{ij}$  y define  $W_{ij}$  como:

$$a_k = \begin{cases} 1 & \text{si (escena } k) \in \Omega, \\ 0 & \text{en los demás casos} \end{cases} \quad (91)$$

$$W_{i,j} = \frac{a_j D_{i,j}^{-1}}{\sum_{k=1}^{12} a_k D_{i,j}^{-1}} \quad (92)$$

Sea ahora  $i = \{1, 2, \dots, 9\} \equiv \{(plane, Y), (plane, C_B), (plane, C_R), (edge, Y), (edge, C_B), (edge, C_R), (texture, Y), (texture, C_B), (texture, C_R)\}$ , siendo  $(plane, C)$ ,  $(edge, C)$  y  $(texture, C)$  las regiones de plana, bordes y textura de la componente de imagen  $C$ ,  $C = Y, C_B, C_R$ .

Sea  $u = texture, edge, plane$  y  $v = Y, C_B, C_R$ , se calcularán los valores  $W_i$ ,  $i = 1, 2, \dots, 9$ , como:

$$E_i = \sum_{j=1}^{12} D_{i,j} W_{i,j}$$

$$\kappa_{u,v} = \begin{cases} 1 & \text{si } v = Y_i \\ \frac{1}{2} & \text{en los demás casos} \end{cases}$$

$$\tau = \sum_u \left[ \frac{1}{E_{u,Y}} + \frac{1}{2} \left( \frac{1}{E_{u,C_B}} + \frac{1}{E_{u,C_R}} \right) \right]$$

$$W_i = \frac{\kappa_i}{\tau} \cdot \frac{1}{E_i} \quad (93)$$

## 7.2 Cálculo de $F_i$ y $G_i$

Los niveles de degradación contextual  $L_i^{(420)}$  y  $L_i^{(CIF)}$  del cuadro  $f$  para CD420 y CDCIF se calculan del siguiente modo:

$$L_i^{(420)} = \frac{1}{\gamma} \cdot \sum_{j=1}^{12} W_{i,j} L_{i,j}^{(420)} \quad (94)$$

$$L_i^{(CIF)} = \frac{1}{\gamma} \cdot \sum_{j=1}^{12} W_{i,j} L_{i,j}^{(CIF)} \quad (95)$$

siendo  $\gamma$  un factor restringido al intervalo  $[1/2, 2]$ , que se calcula a partir de las distancias vectoriales,  $D_j$ , entre los atributos espaciales y temporales (véase el § 5),  $(S_j, T_j)$  y  $(\bar{S}_j, \bar{T}_j)$ , de la escena de entrada y de cada escena de la base de datos, respectivamente.

$$D_j = (S - \bar{S}_j)^2 + (T - \bar{T}_j)^2 \quad (96)$$

$$w_j = \frac{D_j^{-1}}{\sum_{j=1}^{12} D_k^{-1}}$$

$$a = \sum_{j=1}^{12} w_j \left[ \frac{\bar{S}_j \cdot \bar{T}_j}{2} + (1 - \bar{T}_j^2) \cdot \left( 1 - \frac{\bar{S}_j^2}{2} \right) \right] \quad (97)$$

$$b = \frac{ST}{2} + (1 - T^2) \cdot \left( 1 - \frac{S^2}{2} \right)$$

$$\gamma = 1 + a - b$$

Por último, los parámetros  $F_i$  y  $G_i$  se obtienen resolviendo el siguiente sistema de ecuaciones:

$$L_i^{(420)} = 100 / \left[ 1 + \left( \frac{F_i}{m_i^{(420)}} \right)^{G_i} \right] \quad (98)$$

$$L_i^{(CIF)} = 100 / \left[ 1 + \left( \frac{F_i}{m_i^{(CIF)}} \right)^{G_i} \right] \quad (99)$$

## 8 Referencias Bibliográficas

GONZALEZ, R. C. y WOODS, R. E. [1992] *Digital Image Processing*. Addison-Wesley.

ISO/CEI [1992] Norma ISO/CEI 11172 – Information technology – Coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbit/s.

TRUCCO, E. y VERRI, A. [1998] *Introductory Techniques for 3-D Computer Vision*. Prentice-Hall.

## Anexo 4a

## Resultados objetivos de las pruebas VQEG-Fase II

CUADRO 10

## Matriz de datos objetivos brutos para 625/60

SRC	HRC									
	1	2	3	4	5	6	7	8	9	10
1		0,6343	0,5083	0,287		0,2461		0,1951		0,1548
2		0,5483	0,5966	0,3649		0,3185		0,2668		0,1597
3		0,5998	0,6299	0,4551		0,3927		0,3428		0,2553
4		0,6055	0,8159	0,5684		0,5397		0,4158		0,309
5		0,6483	0,7268	0,4358		0,418		0,2874		0,1898
6		0,6146	0,4908	0,3671		0,3139		0,2562		0,2107
7				0,5865		0,5536			0,4841	0,3917
8				0,5023		0,457			0,3949	0,3158
9				0,4563		0,3927			0,3399	0,2667
10				0,706		0,6511			0,6025	0,5083
11	0,8124				0,6374		0,3205			0,3221
12	0,7015				0,547		0,4997			0,3922
13	0,709	0,5098					0,4199			0,3298

CUADRO 11

## Matriz de datos objetivos brutos para 525/60

SRC	HRC													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0,5472	0,3698	0,3429	0,1918										
2	0,5075	0,226	0,1028	0,0789										
3	0,3549	0,127	0,058	0,0339										
4					0,6062	0,419	0,36	0,3108						
5					0,4444	0,2957	0,2152	0,1635						
6					0,6098 <sup>(1)</sup>	0,3462	0,2546	0,1967						
7					0,2404	0,135	0,0864	0,0609						
8									0,8666	0,7554	0,6944	0,7048	0,6685	0,494
9									0,8896	0,7134	0,6204	0,6504	0,6246	0,2326
10									0,8776	0,6419	0,4788	0,6392	0,6237	0,1571
11									0,8623	0,7207	0,5719	0,5619	0,5796	0,3012
12									0,8262	0,6193	0,5139	0,5391	0,4946	0,1992
13									0,8223	0,5609	0,3454	0,437	0,4246	0,215

<sup>(1)</sup> Se ha suprimido el valor SRC = 6, HRC = 5 del análisis por superar los requisitos de registro temporal del plan de la prueba VQEG.

## Anexo 5

### Modelo 4

En este Anexo se presenta una descripción funcional completa de la VQM de la NTIA y sus técnicas de calibración asociadas.

Los algoritmos de calibración descritos en este Anexo garantizan sobradamente el funcionamiento adecuado del estimador de calidad de vídeo de la NTIA. En general, estos algoritmos tienen una precisión de registro espacial de más menos 1/2 píxel y una precisión de registro temporal de más menos un campo entrelazado.

#### ÍNDICE

	<i>Página</i>
1	Introducción ..... 48
2	Referencias normativas ..... 48
3	Definiciones ..... 48
4	Esquema del cálculo de la VQM ..... 52
5	Muestreo ..... 53
5.1	Indexación temporal de los ficheros de vídeo original y procesado ..... 54
5.2	Indexación espacial de los cuadros de vídeo original y procesado ..... 55
5.3	Especificación de las subregiones rectangulares ..... 56
5.4	Consideraciones sobre las secuencias de vídeo de más de 10 s ..... 56
6	Calibración ..... 56
6.1	Registro espacial ..... 57
6.1.1	Sinopsis ..... 57
6.1.2	Cuestiones de entrelazado ..... 59
6.1.3	Entradas requeridas al algoritmo de registro espacial ..... 60
6.1.4	Subalgoritmos utilizados por el algoritmo de registro espacial ..... 61
6.1.5	Registro espacial con escenas arbitrarias ..... 62
6.1.6	Registro espacial del vídeo progresivo ..... 68
6.2	Región válida ..... 69
6.2.1	Algoritmo central de la región válida ..... 69
6.2.2	Aplicación del algoritmo central de región válida a una secuencia de vídeo ..... 70
6.2.3	Observaciones sobre el algoritmo de región válida ..... 71

*Página*

6.3	Ganancia y desplazamiento .....	72
6.3.1	Algoritmo central de ganancia y desplazamiento de nivel .....	72
6.3.2	Utilización de escenas .....	73
6.3.3	Aplicación de las correcciones de la ganancia y del desplazamiento de nivel .....	75
6.4	Registro temporal .....	75
6.4.1	Algoritmo basado en tramas para estimar los retardos temporales variables entre las secuencias de vídeo original y procesada.....	76
6.4.2	Aplicación de la corrección del registro temporal .....	80
7	Características de calidad .....	80
7.1	Introducción .....	80
7.1.1	Regiones S-T .....	81
7.2	Características basadas en gradientes espaciales .....	83
7.2.1	Filtros de realce de bordes .....	83
7.2.2	Descripción de las características $f_{SI13}$ y $f_{HV13}$ .....	84
7.3	Características basadas en la información de crominancia .....	87
7.4	Características basadas en la información de contraste .....	87
7.5	Características basadas en ATI .....	87
7.6	Características basadas en el producto vectorial del contraste y la ATI .....	88
8	Parámetros de calidad .....	88
8.1	Introducción .....	88
8.2	Funciones de comparación .....	89
8.2.1	Tasa de errores y tasa logarítmica .....	89
8.2.2	Distancia euclidiana .....	90
8.3	Funciones de agrupamiento espacial .....	91
8.4	Funciones de agrupamiento temporal .....	91
8.5	Recorte y ajuste no lineal .....	93
8.6	Convenio de denominación de parámetros .....	94
8.6.1	Ejemplos de nombres de parámetros .....	97
9	Modelo general .....	98
10	Referencias Bibliográficas.....	100
	Anexo 5a – Datos objetivos brutos de la VQM de la NTIA.....	100

## 1 Introducción

En este Anexo se presenta una descripción técnica completa del modelo general de la National Telecommunications and Information Administration (NTIA) y de las técnicas de calibración asociadas (por ejemplo, la estimación y corrección del registro temporal, del registro espacial y de los errores de ganancia/desplazamiento). El modelo general es el proponente H en las pruebas de televisión de referencia completa VQEG Fase II. El modelo general se diseñó como VQM polivalente para sistemas de vídeo de una gran diversidad de calidades y velocidades binarias. Se llevaron a cabo pruebas exhaustivas, subjetivas y objetivas, para verificar la calidad de funcionamiento del modelo general, antes de someterlo a la prueba VQEG Fase II. Aunque las pruebas VQEG Fase II sólo evaluaron la calidad de funcionamiento del modelo general para los sistemas de vídeo MPEG-2 y de la Recomendación UIT-T H.263, el modelo general debería funcionar correctamente con cualquier otro tipo de sistema de transmisión y codificación.

Los algoritmos de calibración descritos en este Anexo garantizan sobradamente el funcionamiento del estimador de la calidad de vídeo. En general, estos algoritmos tienen una precisión de registro espacial de más menos 1/2 píxel y una precisión de registro temporal de más menos un campo entrelazado.

El modelo general y las técnicas de calibración automática asociadas se han implementado totalmente en soporte lógico de fácil manejo. Este soporte lógico está disponible para todos los interesados sujeto a un acuerdo de licencia de evaluación gratuito (véase [www.its.blrdoc.gov/n3/video/vqmsoftware.htm](http://www.its.blrdoc.gov/n3/video/vqmsoftware.htm) para más información).

## 2 Referencias normativas

Recomendación UIT-R BT.601– Parámetros de codificación de televisión digital para estudios con formatos de imagen normal 4:3 y de pantalla ancha 16:9.

## 3 Definiciones

**4:2:2:** Formato de muestreo de imágenes  $Y$ ,  $C_b$ ,  $C_r$  en el que los planos de crominancia ( $C_b$  y  $C_r$ ) se muestrean horizontalmente a la mitad de velocidad de muestreo del plano de luminancia ( $Y$ ). Véase la Recomendación UIT-R BT.601 (véase el § 2).

**Abertura de producción:** Retícula de imagen que representa la máxima amplitud posible de la imagen en una norma determinada. La apertura de producción representa la amplitud deseable para la adquisición, generación y procesamiento de la imagen antes de la supresión. En el vídeo muestreado de la Recomendación UIT-R BT.601 la apertura de producción es de 720 píxeles  $\times$  486 líneas en los sistemas de 525 líneas y de 720 píxeles  $\times$  576 líneas en los de 625 líneas [SMPTE, 1995b].

**Campo:** Mitad de un cuadro, conteniendo todas las líneas pares o impares.

**Característica:** Cantidad de información asociada a una subregión espacio-temporal de una secuencia de vídeo (ya sea una secuencia original de vídeo o una secuencia de vídeo procesada) o extraída de aquella.

**Circuito de referencia hipotético (HRC, *hypothetical reference circuit*):** Sistema de vídeo objeto de la prueba tal como un códec o un sistema digital de transmisión de vídeo.

**Códec:** Abreviatura de codificador/decodificador o de compresor/descompresor.



**Coordenadas rectangulares:** Subregión de imagen de forma rectangular contenida totalmente en la abertura de producción y especificada por cuatro coordenadas (superior, izquierda, inferior y derecha). La numeración comienza en cero de modo que el ángulo (superior, izquierdo) de la imagen muestreada es (0,0). Véase el § 5.3.

**Corto:** Representación digital de una escena almacenada en un medio informático.

**Crominancia ( $C$ ,  $C_B$ ,  $C_R$ ):** Porción de la señal de vídeo que transporta predominantemente la información de color ( $C$ ), que puede estar separada a su vez en una señal de diferencia de color azul ( $C_B$ ) y una señal de diferencia de color rojo ( $C_R$ ).

**Cuadro:** Una imagen completa de televisión.

**Cuadros por segundo (FPS, *frames per second*):** Número de cuadros originales por segundo transmitidos por el sistema de vídeo objeto de la prueba. Por ejemplo, un sistema de vídeo NTSC transmite aproximadamente 30 fps.

**Desplazamiento o desplazamiento de nivel:** Sumando aplicado por el circuito de referencia hipotético (HRC) a todos los píxeles de un plano de imagen individual (por ejemplo, luminancia, crominancia). El desplazamiento de la señal de luminancia se conoce comúnmente como brillo.

**Escena:** Secuencia de cuadros de vídeo.

**Formato intermedio común (CIF, *common intermediate format*):** Estructura de muestreo de vídeo utilizada en la teleconferencia de vídeo, en la que el canal de luminancia se muestrea a 352 píxeles por 288 líneas (Recomendación UIT-T H.261 – Códec vídeo para servicios audiovisuales a  $p \times 64$  kbit/s).

**Cuarto del formato intermedio común (QCIF, *quarter common intermediate format*):** Estructura de muestreo de vídeo que se utiliza en la teleconferencia de vídeo y en la que el canal de luminancia se muestrea a 176 píxeles por 144 líneas (Recomendación UIT-T H.261).

**Ganancia:** Factor de escala aplicado por el HRC a todos los píxeles de un plano de imagen individual (por ejemplo, luminancia, crominancia). La ganancia de la señal de luminancia se denomina comúnmente contraste.

**Grupo de Expertos en calidad de vídeo (VQEG, *Video Quality Experts Group*):** Grupo de expertos internacionales en calidad de vídeo que realizan las pruebas de validación de las métricas objetivas de calidad de funcionamiento de vídeo. Los resultados del VQEG se comunican a la Unión Internacional de Telecomunicaciones (UIT) y pueden utilizarse como base para las Recomendaciones internacionales de medición de la calidad de vídeo.

**Grupo de Expertos en imágenes en movimiento (MPEG, *Moving Picture Experts Group*):** Grupo de Trabajo del ISO/CEI que se encarga del desarrollo de normas para la representación codificada de audio y vídeo digital (por ejemplo, MPEG-1, MPEG-2 y MPEG-4).

**H.261:** Abreviatura de la Recomendación UIT-T H.261.

**Incertidumbre ( $U$ , *uncertainty*):** Error estimado (más o menos) del registro temporal tras la tolerancia establecida para la mejor hipótesis de retardo de vídeo del HRC. Véase el § 6.4.

**Información espacial (SI, *spatial information*):** Característica basada en estadísticas extraídas de los gradientes espaciales (es decir de los bordes) de una imagen o escena de vídeo. La referencia (Recomendación UIT-T P.910 – Métodos de evaluación subjetiva de la calidad vídeo para aplicaciones multimedios) contiene una definición de SI basada en estadísticas extraídas de  $3 \times 3$  imágenes filtradas con Sobel [Jain, 1989] mientras que el § 7.2 del presente Anexo presenta una definición de SI basada en estadísticas extraídas de  $13 \times 13$  imágenes mucho más grandes filtradas en los bordes (véase la Fig. 29).

**Información temporal (TI, *temporal information*):** Característica basada en estadísticas extraídas de los gradientes temporales (es decir del movimiento) de una escena de vídeo (Recomendación UIT-T P.910) y el § 7.5 de este Anexo contienen definiciones de la TI basada en estadísticas extraídas de diferencias entre cuadros sencillos.

**Información temporal absoluta (ATI, *absolute temporal information*):** Característica derivada del valor absoluto de las imágenes de información temporal calculada como diferencia entre cuadros correlativos de un corto de vídeo. La ATI cuantifica la cantidad de movimiento de una escena de vídeo. Véase en el § 7.5 su definición matemática exacta.

**Línea con alternancia de fase (PAL, *phase-alternate line*):** Sistema compuesto de vídeo analógico en color de 625 líneas.

**Luminancia (Y):** Porción de la señal de vídeo que transporta predominantemente la información de luminancia (o sea la parte en blanco y negro de la imagen).

**Métrica, modelo o medición de la calidad de vídeo (VQM, *video quality metric, model, or measurement*):** Medición global de la degradación de vídeo (véase VQM de corto, modelo general). La VQM se presenta como un solo número con un intervalo de salida nominal entre cero y uno, correspondiendo el cero a la ausencia de percepción de degradación y el uno a la degradación percibida máxima.

**Modelo general:** Modelo de calidad de vídeo, o VQM, objeto de este Anexo 5. El modelo general se sometió a las pruebas de la Fase II llevadas a cabo por el Grupo de Expertos en Calidad de Vídeo (VQEG). El Informe Final del VQEG Fase-II describe la calidad de funcionamiento del modelo general (véase proponente H)<sup>1</sup>.

**Sistema NTSC (National Television Systems Committee):** Sistema compuesto de vídeo analógico en color de 525 líneas [SMPTE, 1999].

**Nota media de opinión (NMO) (en inglés: MOS):** Juicio medio de calidad subjetiva otorgado por un grupo especial de espectadores a un corto de vídeo procesado.

**Parámetro:** Medida de la distorsión de vídeo obtenida de la comparación de dos secuencias paralelas de características, la secuencia de vídeo original y la secuencia correspondiente de vídeo procesado.

**Recomendación UIT-R BT.601:** Normas (véase el § 2) de muestreo de vídeo común de 8 bits para el canal de luminancia (Y) a 13,5 MHz, y los canales de diferencia de color azul y rojo ( $C_B$  y  $C_R$ ) a 6,75 MHz. Véase el § 5 para más información.

**Reencuadre (*reframing*):** Proceso de ordenar dos campos entrelazados de vídeo procesado muestreados correlativamente para formar un cuadro de vídeo. El reencuadre es necesario cuando los HRC no conservan los tipos de campos de entrelazado estándar (por ejemplo, un campo NTSC de tipo uno sale como un campo NTSC de tipo dos y viceversa). Véase el § 6.1.2.

**Referencia reducida:** Metodología de medición de la calidad de vídeo que utiliza características de anchura de banda baja, extraídas de las secuencias de vídeo original o procesada, en contraposición a la utilización de vídeo de referencia completa que requiere el conocimiento completo de las secuencias de vídeo original y procesada (Recomendación UIT-T J.143 – Requisitos de usuario para mediciones objetivas de la percepción de la calidad vídeo en televisión digital por cable). Las metodologías de referencia reducida tienen ventajas para la supervisión en servicio de la calidad extremo a extremo ya que la información de referencia reducida se transmite con facilidad a cualquier punto de la red de telecomunicaciones.

**Región de interés (ROI, *region of interest*):** Réticula de imagen (definida por sus coordenadas rectangulares) utilizada para indicar una subregión particular de un campo o cuadro de vídeo. Véase asimismo SROI.

**Región de interés procesada (PROI, *processed region of interest*):** ROI extraída del vídeo procesado y con los desplazamientos espaciales del HRC corregidos, definida por sus coordenadas rectangulares.

**Región espacial de interés (SROI, *spatial region of interest*):** Réticula de imagen específica (definida por sus coordenadas rectangulares) utilizada para calcular la VQM de un corto de vídeo. La SROI es un subconjunto rectangular que queda totalmente encerrado en la región válida procesada. Para el vídeo muestreado con arreglo a la Recomendación UIT-R BT.601, la SROI recomendada es de 672 píxeles  $\times$  448 líneas para los sistemas de 525 líneas y de 672 píxeles  $\times$  544 líneas para los sistemas de 625 líneas, centradas en la abertura de producción. Esta SROI recomendada corresponde aproximadamente a la porción de imagen de vídeo visible en una pantalla, excluida la zona de sobreexploración. Véase asimismo la ROI.

**Región original de interés (OROI, *original region of interest*):** ROI extraída del vídeo original, definida por sus coordenadas rectangulares.

**Región temporal de interés (TROI, *temporal region of interest*):** Segmento de tiempo específico, secuencia o subconjunto de cuadros utilizados para calcular la VQM de un corto. La TROI es un segmento contiguo de cuadros que queda totalmente encerrado en la región válida temporal. La máxima TROI posible es el segmento temporal completamente registrado y contiene todos los cuadros temporalmente registrados de la TVR. Cuando sea necesario el reencuadre, el corto procesado será el que se reencuadra siempre y no el corto original.

**Región válida (VR, *valid region*):** Porción rectangular de una retícula de imagen (definida por sus coordenadas rectangulares) no suprimida ni corrompida por el procesamiento. La región válida es un subconjunto de la abertura de producción de la norma de vídeo e incluye únicamente aquellos píxeles de imagen que contienen información de imágenes que no hayan sido suprimidas ni corrompidas. Véase región válida original y región válida procesada.

**Región válida original (OVR, *original valid region*):** Región válida de un corto de vídeo original, definida por sus coordenadas rectangulares.

**Región válida procesada (PVR, *processed valid region*):** Región válida de un corto de vídeo procesado por el HRC; definida por sus coordenadas rectangulares. La PVR siempre se define con referencia al vídeo original, de modo que es necesario que el HRC corrija los eventuales desplazamientos espaciales del vídeo antes de calcular la PVR. Por consiguiente, la PVR siempre está contenida en la OVR. La región entre la PVR y la OVR es la porción de vídeo borrada o corrompida por el HRC.

**Región válida temporal (TVR, *temporal valid region*):** Máximo segmento de tiempo, secuencia o subconjunto de cuadros de vídeo que puede utilizarse para el calibrado y cálculo de la VQM. Los cuadros que caen fuera de este segmento de tiempo se considerarán siempre no válidos.

**Registro espacial (*spatial registration*):** Proceso utilizado para estimar y corregir los desplazamientos espaciales de la secuencia de vídeo procesada con respecto a la secuencia de vídeo original.

**Registro temporal:** Proceso utilizado para estimar y corregir el desplazamiento temporal (es decir el retardo de vídeo) de la secuencia de vídeo procesada con respecto a la secuencia de vídeo original (véase el § 6.4.1).

**Sobreexploración:** Porción de vídeo que no suele ser visible en una pantalla de televisión normal.

**SMPTE (Society of Motion Picture and Television Engineers):** Institución líder en la industria del cine y de la televisión, dedicada al desarrollo de la teoría y de la aplicación de las imágenes en movimiento, entre ellas las de cine, televisión, vídeo, computadores y telecomunicaciones. La industria confía a la SMPTE la generación de normas, directrices de ingeniería y prácticas recomendadas destinadas a los respectivos campos profesionales.

**Subregión espacial-temporal (S-T) (*spatial-temporal (S-T) subregion*):** Bloque de píxeles de imagen de una secuencia de vídeo original o procesado que incluye un tramo vertical (número de filas), un tramo horizontal (número de columnas), y un tramo temporal (número de cuadros). Véase la Fig. 27.

**Unidad del Institute for Radio Engineers (IRE):** Unidad de tensión que se utiliza comúnmente para la medición de señales de vídeo. Un IRE es equivalente a 1/140 de voltio.

**Unión Internacional de Telecomunicaciones (UIT):** Organización Internacional del sistema de Naciones Unidas en la que los gobiernos y el sector privado coordinan las redes y servicios de telecomunicaciones mundiales. La UIT comprende el Sector de Radiocomunicaciones (UIT-R) y el Sector de Normalización de las Telecomunicaciones (UIT-T).

**Vídeo de entrada:** Vídeo antes de ser procesado o distorsionado por un HRC (véase la Fig. 19). El vídeo de entrada puede denominarse también vídeo original.

**Vídeo de salida:** Vídeo procesado o distorsionado por el HRC (véase la Fig. 19). El vídeo de salida recibe también el nombre de vídeo procesado.

**Vídeo original:** Vídeo antes de ser procesado o distorsionado por el HRC (véase la Fig. 19). El vídeo original recibe además el nombre de vídeo de entrada por ser la entrada de vídeo al sistema digital de transmisión de vídeo.

**Vídeo procesado:** Vídeo procesado o distorsionado por el HRC (véase la Fig. 19). El vídeo procesado recibe también el nombre de vídeo de salida por ser la salida de vídeo del sistema digital de transmisión de vídeo.

**VQM de corto:** VQM de un solo corto de vídeo procesado.

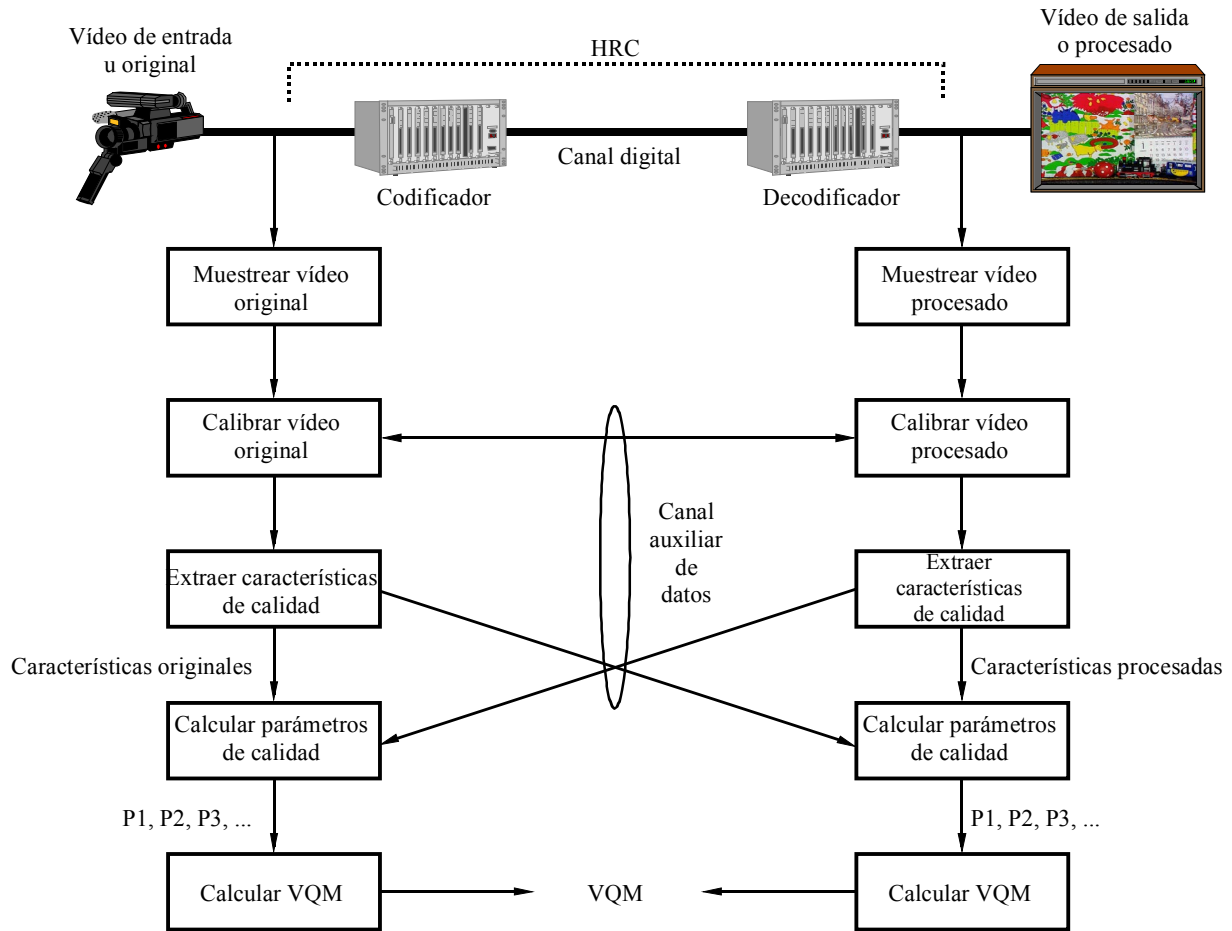
**YUV grande:** Formato de fichero binario utilizado para almacenar cortos que se hayan muestreado con arreglo a la Recomendación UIT-R BT.601. En el formato *YUV* grande, todos los cuadros de vídeo de una escena se almacenan en un fichero binario grande, en el que cada uno de los cuadros se ajusta al muestreo de la Recomendación UIT-R BT.601. *Y* representa la información del canal de luminancia, *U* representa el canal de diferencia de color azul (es decir  $C_B$  en la Recomendación UIT-R BT.601), y *V* representa el canal de diferencia de color rojo (o sea  $C_R$  en la Recomendación UIT-R BT.601). El orden de los píxeles en el fichero binario coincide con el especificado en SMPTE 125M [SMPTE, 1995a]. La especificación completa del formato de fichero *YUV* grande figura en el § 5 mientras que las rutinas del soporte lógico para la lectura y visualización de los ficheros *YUV* grandes figura en [Pinson y Wolf, 2002].

#### 4 Esquema del cálculo de la VQM

Este Anexo presenta una descripción completa del modelo general y de los algoritmos de calibración asociados. Estos algoritmos de medición objetiva automática proporcionan buenas aproximaciones de las impresiones de calidad global o notas medias de opinión, de las degradaciones de vídeo digital sometidas a la calificación de grupos especiales de espectadores (Recomendación UIT-R BT.500). La Fig. 19 proporciona un diagrama esquemático de los procesos necesarios para calcular la VQM general. Entre estos procesos se encuentra el muestreo de las secuencias de vídeo original y procesada (véase el § 5), la calibración de las secuencias de vídeo

original y procesada (véase el § 6), la extracción de las características basadas en la percepción (véase el § 7), el cálculo de los parámetros de calidad del vídeo (véase el § 8), y el cálculo del modelo general (véase el § 9). El modelo general controla el estado de los cambios perceptuales de calidad debidos a las distorsiones en cualquier componente del sistema digital de transmisión de vídeo (por ejemplo, el codificador, la cadena digital, el decodificador).

FIGURA 19  
Pasos necesarios para calcular la VQM



1683-19

El método de medición presentado en este documento utiliza parámetros de referencia reducida de anchura de banda alta (Recomendación UIT-T J.143). Estos parámetros de referencia reducida utilizan características extraídas de las regiones espacio-temporales (S-T) de la secuencia de vídeo (véase el § 7.1.1). Por consiguiente, el método de medición presentado aquí puede utilizarse también para realizar la supervisión en servicio de la calidad de vídeo en situaciones en las que se dispone de un canal de datos auxiliar para transmitir las características extraídas entre los extremos fuente y destino de un HRC como se representa en la Fig. 19.

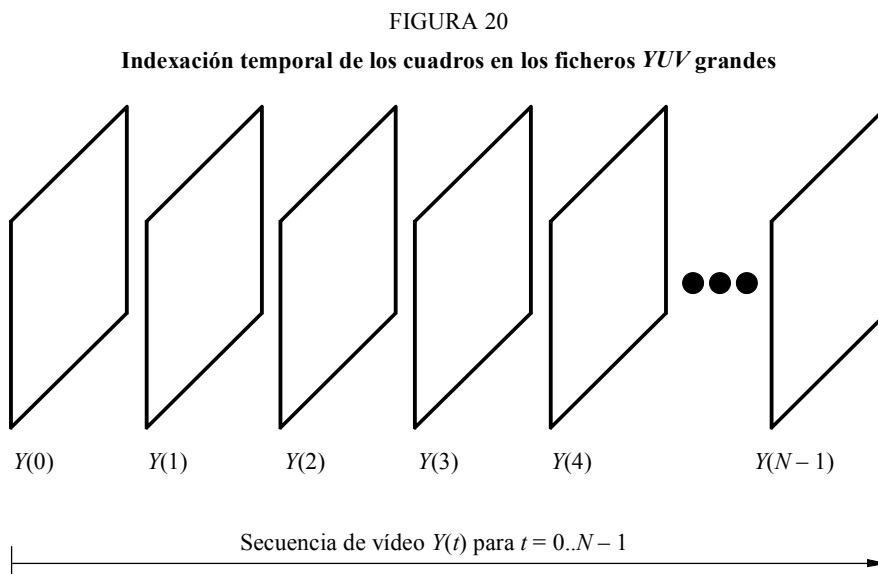
## 5 Muestreo

Los algoritmos de este Anexo se han diseñado para ser ejecutados en un computador bajo la hipótesis de que las secuencias de vídeo original y procesada estén disponibles en forma de representaciones digitales almacenadas en un medio informático (denominadas cortos en este Anexo).

Para el vídeo en formato analógico, una de las normas de muestreo digital más utilizadas es la Recomendación UIT-R BT.601 (véase el § 2). El vídeo compuesto, tal como el NTSC o el PAL, debe convertirse antes en vídeo de componentes que contenga las tres señales siguientes: luminancia ( $Y$ ), diferencia de color azul,  $C_B$ , y diferencia de color rojo,  $C_R$ . El muestreo de la Recomendación UIT-R BT.601 también suele denominarse comúnmente muestreo 4:2:2, ya que el canal  $Y$  se muestrea a velocidad completa mientras que los canales  $C_B$  y  $C_R$  se muestrean a media velocidad. La Recomendación UIT-R BT.601 especifica una velocidad de muestreo de 13,5 MHz que produce 720 muestras  $Y$  por línea de vídeo. Dado que en la norma NTSC de 525 líneas hay 486 líneas que contienen información de imagen, el cuadro completo  $Y$  muestreado con arreglo a la Recomendación UIT-R BT.601 tendrá 720 píxeles por 486 líneas. Análogamente, cuando se muestrea vídeo PAL de 625 líneas con arreglo a la Recomendación UIT-R BT.601, el cuadro de vídeo  $Y$  contendrá 720 píxeles por 576 líneas. Si se utilizan 8 bits para muestrear uniformemente la señal  $Y$ , la Recomendación UIT-R BT.601 especifica que el negro de referencia (es decir 7,5 unidades IRE) se muestrearán como «16» y que el blanco de referencia (es decir 100 unidades IRE) se muestrearán como «235». Por consiguiente, hay un margen de trabajo disponible para las señales de vídeo que sobrepasen los niveles de referencia blanco y negro antes de que sean recortados por el convertidor analógico-digital. Cada uno de los canales de crominancia ( $C_B$  y  $C_R$ ) se muestrea a 6,75 MHz, de modo que el primer par de muestras de crominancia ( $C_B$ ,  $C_R$ ) se asocia a la primera muestra de luminancia  $Y$ , el segundo par de muestras de crominancia se asocia a la tercera muestra de luminancia, y así sucesivamente. Dado que los canales de crominancia son bipolares, la señal cero se muestrea como «128».

### 5.1 Indexación temporal de los ficheros de vídeo original y procesado

Un cuadro de vídeo de luminancia obtenido del muestreo con arreglo a la Recomendación UIT-R BT.601, se representará por  $Y(t)$ . La variable  $t$  se utiliza aquí como índice para direccionar los cuadros muestreados de los ficheros original y procesado de  $YUV$  grande, pero no indica tiempo real. Si el fichero de  $YUV$  grande contiene  $N$  cuadros, como se representa en la Fig. 20,  $t = 0$  indica el primer cuadro muestreado mientras que  $t = (N - 1)$  indica el último.



Todos los algoritmos se desarrollan y describen para su ejecución con pares de ficheros muestreados: una secuencia de vídeo original y su secuencia asociada de vídeo procesado. Para evitar confusiones, se supone que ambos ficheros tienen la misma longitud. Además, se establece la hipótesis inicial de que el primer cuadro del fichero original se alinea temporalmente con el primer cuadro del fichero procesado, con más o menos cierta incertidumbre temporal.

En las implementaciones en servicio en tiempo real, esta hipótesis de incertidumbre equilibrada puede sustituirse por una incertidumbre unilateral. La causalidad limita el rango de la incertidumbre temporal. Por ejemplo, un cuadro procesado que ocurra en el instante  $t = n$  debe proceder de cuadros originales que ocurran en un instante igual o anterior a  $t = n$ .

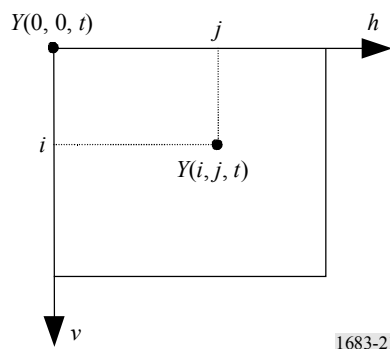
La hipótesis anterior relativa a los ficheros de vídeo original y procesado (es decir la alineación de sus primeros cuadros) equivale a seleccionar la mejor suposición para el retardo temporal del HRC representado en la Fig. 19. Por consiguiente, la incertidumbre restante de la estimación de retardo de vídeo se representará por más o menos  $U$ .

## 5.2 Indexación espacial de los cuadros de vídeo original y procesado

En la Fig. 21 se representa el sistema de coordenadas utilizado en los cuadros de luminancia muestreados. Las coordenadas horizontal y vertical del vértice superior izquierda del cuadro de luminancia se definen como  $(v = 0, h = 0)$ , de modo que la coordenada del eje horizontal,  $h$ , aumenta hacia la derecha y la del eje vertical,  $v$ , hacia abajo. Las coordenadas del eje horizontal van de cero a uno menos el número de píxeles de la línea. Las coordenadas del eje vertical van de cero a uno menos el número de líneas de la imagen, lo que se especificará en líneas de cuadro para los sistemas progresivos y en líneas de campo o líneas de cuadro para los sistemas entrelazados. La amplitud de un píxel muestreado en  $Y(t)$  en la fila  $i$  (o sea para  $v = i$ ), columna  $j$  (o sea para  $h = j$ ), y tiempo  $t$  se representa por  $Y(i, j, t)$ .

FIGURA 21

Sistema de coordenadas utilizado en los cuadros de luminancia muestreada  $Y$

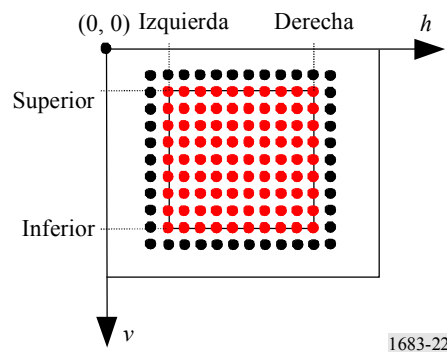


Un corto de vídeo muestreado con arreglo a la Recomendación UIT-R BT.601 se guarda en el formato de fichero «YUV grande», donde  $Y$  representa la información de luminancia de la Recomendación UIT-R BT.601,  $U$  representa la información de diferencia de color azul (es decir  $C_B$  en la Recomendación UIT-R BT.601), y  $V$  representa la información de diferencia de color rojo (es decir  $C_R$  en la Recomendación UIT-R BT.601). En el formato de fichero YUV grande, todos los cuadros se almacenan secuencialmente en un gran fichero binario continuo. Los píxeles de imagen se almacenan secuencialmente por líneas de exploración de vídeo como bytes en el orden siguiente:  $C_{B0}, Y_0, C_{R0}, Y_1, C_{B2}, Y_2, C_{R2}, Y_3$ , etc., donde el subíndice numérico representa el número de píxel (debe utilizarse la replicación o interpolación de píxeles para encontrar las muestras de crominancia  $C_B$  y  $C_R$  de  $Y_1, Y_3, \dots$ ). Este orden de bytes es equivalente al especificado en SMPTE 125M [SMPTE, 1995a].

### 5.3 Especificación de las subregiones rectangulares

En una imagen muestreada se utilizan subregiones rectangulares para controlar el cálculo de la VQM. Por ejemplo, la VQM puede calcularse en una región válida de la imagen muestreada o en una región espacial de interés especificada por el usuario que sea más pequeña que la región válida. La especificación de subregiones rectangulares utilizará coordenadas rectangulares definidas por las cuatro magnitudes superior, izquierda, inferior, y derecha. En la Fig. 22 se representa la especificación de una subregión rectangular correspondiente a un solo cuadro de vídeo muestreado. Los píxeles rojos de la imagen están incluidos en la subregión pero los píxeles negros de la imagen están excluidos. En el cálculo de la VQM, se suele dividir una imagen en un gran número de pequeñas subregiones contiguas. La definición de subregión rectangular utilizada en la Fig. 22 permite definir la rejilla utilizada para mostrar estas subregiones contiguas y los cálculos realizados para extraer las características de cada subregión contigua.

FIGURA 22  
Coordenadas rectangulares para la especificación de subregiones de imagen



### 5.4 Consideraciones sobre las secuencias de vídeo de más de 10 s

Las mediciones de calidad de vídeo del presente Anexo se basaron en los resultados de pruebas subjetivas en las que se utilizaron cortos de vídeo de 8 a 10 s. Cuando se trabaja con secuencias de vídeo de mayor duración, deben dividirse éstas en segmentos de vídeo más cortos, en los que se supone que cada segmento tiene su propia calibración y atributos de calidad. La división de la secuencia de vídeo en segmentos solapados y el procesamiento independiente de cada segmento es uno de los métodos de emular la evaluación de calidad continua de secuencias de vídeo largas por medio de las técnicas VQM aquí presentadas.

## 6 Calibración

Se necesitan cuatro Pasos para calibrar adecuadamente el vídeo muestreado a fin de prepararlo para la extracción de sus características. Estos Pasos son los siguientes:

*Paso 1:* Estimación y corrección del registro espacial.

*Paso 2:* Estimación de la región válida para limitar la extracción de características a aquellos píxeles que contengan información de la imagen.

*Paso 3:* Estimación y corrección del desplazamiento de nivel y ganancia (normalmente conocidos como contraste y brillo).

*Paso 4:* Estimación y corrección del registro temporal.



El Paso 2 debe ejecutarse en la secuencia de vídeo original y también en la procesada. Los Pasos 1, 3 y 4 deben ejecutarse en la secuencia de vídeo procesada. Normalmente, el registro espacial, la ganancia y el desplazamiento de nivel son constantes para un sistema de vídeo determinado y por consiguiente sólo es necesario calcular estos valores una vez. No obstante, suele ser común que la región válida y el registro temporal cambien en función del contenido de la escena. Por ejemplo, las escenas de pantalla completa y las que dejan bandas negras en la parte superior e inferior de la pantalla tendrán distintas regiones válidas, y los sistemas de videoconferencia suelen tener retardos de vídeo variables en función del contenido de las escenas (por ejemplo, orador en contraste con escena deportiva). Además de las técnicas de calibración aquí presentadas, el lector puede desear examinar (Recomendación UIT-T P.931 – Retardo de las comunicaciones multimedia, sincronización y medición de la velocidad de tramas) que contiene métodos alternativos de registro espacial y temporal.

La calibración previa a la extracción de características significa que la VQM no será sensible a los desplazamientos horizontales y verticales de la imagen, ni a los desplazamientos temporales de la secuencia de vídeo provocados por retardos no nulos de vídeo, ni a los cambios de contraste y brillo de la imagen que caigan dentro del margen dinámico de la unidad de muestreo de vídeo. Aunque estas magnitudes de calibración tengan una repercusión significativa en la calidad global percibida (por ejemplo, imágenes de bajo contraste de un sistema de vídeo con una ganancia de 0,3), la filosofía que se ha seguido aquí consiste en comunicar la información de calibración independientemente de la VQM. Los desplazamientos espaciales, las regiones válidas, las ganancias y los desplazamientos suelen poder ajustarse por medio de buenas prácticas de ingeniería, mientras que los retardos temporales proporcionan importante información de calidad cuando se evalúan sistemas de vídeo bidireccionales o interactivos.

Todos los parámetros y características de calidad de vídeo (véanse los § 7 y 8) suponen que sólo se suprimirá un retardo de vídeo para registrar temporalmente la secuencia de vídeo procesada (o sea, retardo de vídeo constante). Algunos sistemas de vídeo o HRC retardan cuadros procesados particulares en distintas magnitudes (es decir, con retardo de vídeo variable). A los efectos del presente Anexo, todos los sistemas de vídeo se tratan como si tuvieran retardo de vídeo constante. Las variaciones con respecto a este retardo se consideran degradaciones que se miden por las características y los parámetros. Esta solución parece producir correlaciones más altas con respecto a la nota subjetiva que las mediciones de calidad de vídeo basadas en secuencias de vídeo procesado en las que se ha suprimido el retardo de vídeo variable. Cuando se trabaja con secuencias de vídeo largas (véase el § 5.4), éstas deben dividirse en segmentos de vídeo más cortos, de modo que cada segmento tenga su propio retardo de vídeo constante. Esto permite cierta variación del retardo en función del tiempo. Puede obtenerse una estimación más continua de las variaciones del retardo dividiendo la secuencia en segmentos de tiempo solapados.

Si además el HRC objeto de la prueba ajusta espacialmente la imagen o modifica su tamaño (por ejemplo, mediante zum), habría que incluir un paso adicional para estimar y suprimir este ajuste espacial en el proceso de calibración. El ajuste espacial va más allá del propósito del presente Anexo.

## **6.1 Registro espacial**

### **6.1.1 Sinopsis**

El proceso de registro espacial determina el desplazamiento espacial horizontal y vertical del vídeo procesado respecto al vídeo original. Un desplazamiento horizontal positivo se asocia a una imagen procesada que se haya movido hacia la derecha dicho número de píxeles. Un desplazamiento vertical positivo se asocia a una imagen procesada que se haya movido hacia abajo dicho número de líneas. Así pues, el registro espacial del vídeo entrelazado genera tres números: el desplazamiento horizontal en píxeles, el desplazamiento del campo vertical uno en líneas de campo, y el

desplazamiento del campo vertical dos en líneas de campo. El registro espacial del vídeo progresivo genera dos números: el desplazamiento horizontal y el desplazamiento vertical en líneas de cuadro. La precisión del algoritmo de registro espacial es al píxel más próximo para los desplazamientos horizontales y a la línea más próxima para los desplazamientos verticales. Una vez calculado el registro espacial, se elimina el desplazamiento espacial de la secuencia vídeo procesado (por ejemplo, una imagen, procesada que se hubiera desplazado hacia abajo se desplaza de nuevo hacia arriba). En el vídeo entrelazado, esto puede incluir el reencuadre de la secuencia de vídeo procesada derivado de la comparación de los desplazamientos de los campos verticales uno y dos.

Cuando se trabaja con vídeo entrelazado, todas las operaciones consideran el vídeo de cada campo por separado; cuando se trabaja con vídeo progresivo, todas las operaciones consideran todo el cuadro de vídeo simultáneamente. Para mayor simplicidad, el algoritmo de registro espacial se describirá completamente, en primer lugar, para el vídeo entrelazado, ya que éste es el caso más complicado. Las modificaciones necesarias para trabajar con vídeo progresivo se definen en el § 6.1.6.

El registro espacial debe determinarse antes que la PVR, la ganancia y el desplazamiento de nivel, y el registro temporal. En particular, cada uno de estos valores debe calcularse por comparación entre los contenidos de vídeo original y procesado que se hayan registrado espacialmente. Si la secuencia de vídeo procesada se hubiera desplazado espacialmente con respecto a la secuencia de vídeo original, no habiéndose corregido este desplazamiento espacial, estos cálculos estarían corrompidos porque se habrían basado en contenidos de vídeo dispares. Por desgracia, el registro espacial no puede determinarse correctamente salvo que se conozcan también la PVR, la ganancia y el desplazamiento de nivel, y el registro temporal. La interdependencia de estos valores provoca un problema de medición «del huevo o la gallina». El cálculo del registro espacial para un campo procesado exige el conocimiento de la PVR, la ganancia y el desplazamiento de nivel, y el campo original que más se ajusta. Pero estos valores no se pueden determinar hasta que no se conozca el desplazamiento espacial. Si hubiera una gran incertidumbre en los valores anteriores, la búsqueda exhaustiva de todas las variables exigiría un enorme número de cálculos.

La solución aquí presentada ejecuta una búsqueda iterativa del campo original que más se ajusta a cada campo procesado. En esta búsqueda se incluye la actualización iterativa de las estimaciones de la PVR, la ganancia y el desplazamiento de nivel, y el registro temporal. En ciertos campos procesados, no obstante, puede fallar el algoritmo de registro espacial. Normalmente, cuando se estima incorrectamente el registro espacial para un campo procesado, la ambigüedad se debe a las características de la escena. Considérese, por ejemplo, una escena entrelazada creada digitalmente, que contenga una toma panorámica hacia la izquierda. Debido a que la toma panorámica se ha generado por computador, esta escena podría tener una toma panorámica horizontal de un píxel por campo, exactamente. Para el algoritmo de búsqueda del registro espacial, sería imposible distinguir entre el registro espacial correcto calculado a partir del campo original emparejado, y un desplazamiento horizontal de dos píxeles calculado mediante el campo que está dos campos antes del campo original emparejado. Otro ejemplo sería el siguiente, considérese una imagen consistente únicamente en líneas verticales blancas y negras digitalmente perfectas. Como la imagen no contiene líneas horizontales, el desplazamiento vertical es totalmente ambiguo. Como el patrón de líneas verticales se repite, el desplazamiento horizontal es ambiguo, ya que dos o más desplazamientos horizontales son igualmente probables.

Por consiguiente, el algoritmo de búsqueda iterativa debe aplicarse a una secuencia de campos procesados. Las estimaciones individuales de los desplazamientos espaciales de varios campos procesados pueden utilizarse entonces para producir una estimación más ajustada. Las estimaciones de desplazamiento espacial de varias secuencias o escenas pueden combinarse a su vez para producir una estimación más ajustada del HRC objeto de la prueba; suponiendo que el desplazamiento espacial sea constante para todas las escenas que pasan por el HRC.

### 6.1.2 Cuestiones de entrelazado

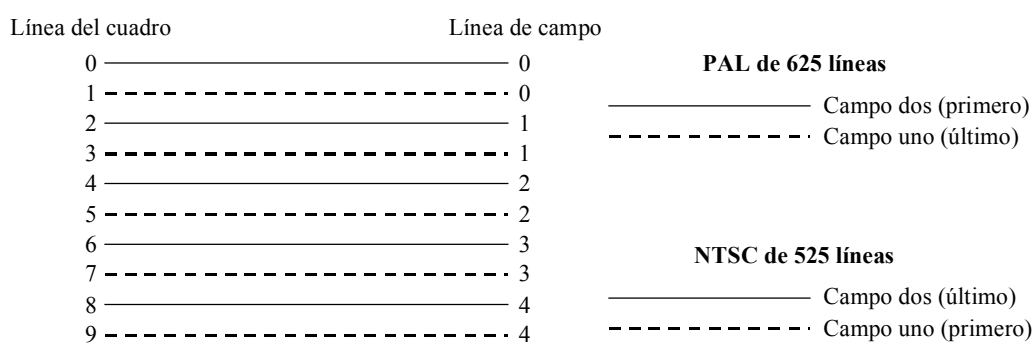
El registro espacial vertical del vídeo entrelazado constituye un problema mayor que el del vídeo progresivo, ya que el proceso de registro espacial debe distinguir entre el campo uno y el campo dos. Hay tres condiciones de desplazamiento vertical que deben distinguirse para obtener un registro correcto de desplazamiento vertical en los sistemas entrelazados: que el campo vertical uno sea igual al campo vertical dos, que el campo vertical uno sea menor que el campo vertical dos, y cualquier otro caso.

Algunos HRC desplazan el campo uno y el campo dos por igual, produciendo un desplazamiento del campo vertical uno igual al desplazamiento del campo vertical dos. En los HRC que no repiten campos ni cuadros (es decir los HRC que transmiten la velocidad de cuadro completa de la norma de vídeo), esta condición supone que lo que era un campo uno en la secuencia de vídeo original es también un campo uno en la secuencia de vídeo procesada y que lo que era un campo dos en el original es asimismo un campo dos en el procesado.

Otros HRC reencuadran el vídeo, desplazando el cuadro muestreado en un número impar de líneas de cuadro. Lo que era campo uno del original se convierte en campo dos de lo procesado, y lo que era campo dos del original se convierte en el campo uno del cuadro siguiente. Visualmente, el vídeo en pantalla parece correcto debido a que la vista no puede percibir un desplazamiento de cuadro de una línea en el vídeo.

Como muestra la Fig. 23, el campo uno comienza en la línea uno del cuadro, y contiene todas las líneas impares del cuadro. El campo dos comienza en la línea cero del cuadro (línea superior del cuadro), y contiene todas las líneas pares del cuadro. En NTCS, el campo uno se presenta en un instante anterior mientras que el campo dos se presenta en un instante posterior. En PAL, el campo dos se presenta en un instante anterior mientras que el campo uno se presenta en un instante posterior.

FIGURA 23  
Diagrama que muestra los campos entrelazados y el esquema de numeración de líneas de cuadro/campo



El reencuadre tiene lugar cuando o bien el primer campo se traslada al último campo o cuando el último campo se traslada al primer campo del siguiente cuadro (retardo de un campo), o bien cuando el último campo se traslada al primer campo y el primer campo del siguiente cuadro se traslada al último campo del cuadro actual (adelanto de un campo). Por ejemplo, cuando el campo original en NTSC se traslada al siguiente campo uno del cuadro NTSC, la línea superior del cuadro se traslada de la línea cero del cuadro del campo dos original a la línea uno del cuadro del campo uno procesado. En la numeración de líneas de campo, la línea superior permanece en la línea de

campo cero, de modo que el campo uno procesado tiene un desplazamiento vertical nulo (ya que los desplazamientos verticales se miden para cada campo en líneas de campo). Cuando el campo uno NTSC original se traslada al campo dos de dicho cuadro, la línea superior del campo se traslada del campo uno original, línea uno del cuadro, al campo dos procesado, línea dos del cuadro. En la numeración de líneas de campo, la línea superior se traslada de la línea de campo cero a la línea de campo uno, de modo que el campo dos procesado tiene un desplazamiento vertical de una línea de campo. La regla general, tanto para NTSC como para PAL, es que cuando el desplazamiento vertical del campo dos (en líneas de campo) sea uno mayor que el desplazamiento vertical del campo uno (en líneas de campo) se ha producido un reencuadre.

Si el desplazamiento vertical del campo dos no es igual al desplazamiento vertical del campo uno, o a uno más, el HRC ha corrompido el muestreo espacial correcto de los dos campos entrelazados de vídeo y el vídeo resultante parecerá oscilar hacia arriba y hacia abajo. Esta degradación es a la vez evidente y molesta para el observador, y por consiguiente no se suele producir en la práctica debido a que cuando el diseñador del HRC descubre este error lo corrige. Por consiguiente, la mayor parte del tiempo, el registro espacial se simplifica en dos patrones comunes. En los sistemas sin reencuadre, el desplazamiento vertical del campo uno es igual al desplazamiento vertical del campo dos; mientras que en los sistemas con reencuadre, el desplazamiento vertical del campo uno más uno es igual al desplazamiento vertical del campo dos.

Obsérvese, además, que el registro espacial incluye cierta información de registro temporal, en particular si el vídeo se ha reencuadrado o no. El proceso de registro temporal puede o no detectar el reencuadre, pero aun pudiendo, el reencuadre es intrínseco del proceso de registro espacial. Por consiguiente, el registro espacial debe poder determinar si el campo procesado que se examina se alinea mejor con un campo uno original o con un campo dos. El registro espacial correspondiente a cada campo sólo puede calcularse correctamente cuando el campo procesado se compara con el campo original que lo creó. Además de la cuestión del reencuadre, la utilización de un campo original equivocado (campo uno en vez de campo dos) puede producir imprecisiones de registro espacial debidas a las diferencias intrínsecas en el contenido espacial de los dos campos entrelazados.

### **6.1.3 Entradas requeridas al algoritmo de registro espacial**

Este apartado proporciona una lista de las variables de entrada requeridas por el algoritmo de registro espacial. Estas entradas especifican elementos tales como intervalo de desplazamiento espacial y los campos temporales en los que efectuar la búsqueda. Si estos intervalos están sobredimensionados, la velocidad de convergencia del algoritmo de búsqueda iterativa utilizado para encontrar el desplazamiento espacial puede ser lenta y aumentar la probabilidad de registro espacial falso para escenas con contenido repetitivo (por ejemplo, cuando alguien saluda con la mano). Por contra, si estos intervalos son demasiado restrictivos el algoritmo de búsqueda encontrará, y ampliará lentamente, los límites del intervalo de búsqueda en las sucesivas iteraciones. Aunque esta inteligencia integrada de búsqueda es útil si el usuario se equivoca en las incertidumbres de búsqueda en una pequeña cantidad, el efecto secundario indeseado es el importante aumento del tiempo de ejecución cuando el usuario se equivoca en una gran cantidad. Otra posibilidad es que, en este caso, el algoritmo de búsqueda sea incapaz de encontrar el desplazamiento espacial correcto.

#### **6.1.3.1 Intervalo esperado de desplazamiento espacial**

El intervalo esperado de desplazamiento espacial para vídeos de 525 líneas y de 625 líneas muestreados con arreglo a la Recomendación UIT-R BT.601 está comprendido entre  $\pm 20$  píxeles horizontalmente y  $\pm 12$  líneas de campo verticalmente. Este intervalo esperado de desplazamiento se ha determinado empíricamente procesando datos de vídeo de cientos de HRC. El intervalo esperado de desplazamientos espaciales para vídeo muestreado con arreglo a otros formatos inferiores a los

de la Recomendación UIT-R BT.601 (por ejemplo, CIF), se supone la mitad del observado para los sistemas de 525 líneas y de 625 líneas. Este algoritmo de búsqueda debería funcionar correctamente, aunque con más lentitud, cuando el campo procesado tiene desplazamientos espaciales que caen fuera del intervalo esperado de desplazamiento espacial. Esto se debe a que el algoritmo de búsqueda la expandirá más allá del intervalo esperado de desplazamiento espacial cuando esté justificado. No obstante, si hay excursiones superiores al 50% del intervalo esperado, puede comunicarse la imposibilidad de encontrar el registro espacial correcto.

### **6.1.3.2 Incertidumbre temporal**

El usuario debe especificar asimismo la incertidumbre del registro temporal, es decir, el intervalo de campos originales que ha de examinarse por cada campo procesado. Esta incertidumbre temporal se expresa en número de cuadros antes y después del registro temporal por defecto. Si las secuencias de vídeo original y procesada se almacenan como ficheros, un registro temporal por defecto razonable sería suponer que los primeros cuadros de cada fichero están alineados. La incertidumbre temporal especificada debe ser suficientemente grande para incluir el registro temporal real. Una incertidumbre de más menos un segundo (30 cuadros en el vídeo NTSC de 525 líneas y 25 cuadros en el vídeo PAL de 625 líneas) debe ser suficiente para la mayor parte de los sistemas de vídeo. Los HRC con grandes retardos de vídeo pueden exigir una incertidumbre temporal mayor. El algoritmo de búsqueda puede examinar registros temporales más allá del intervalo de incertidumbre especificado cuando esté justificado (por ejemplo, cuando el campo original más lejano se escoja como mejor registro temporal).

### **6.1.3.3 Hipótesis de la PVR**

La hipótesis de la PVR especifica la porción de la imagen procesada que no ha sido suprimida ni corrompida en el proceso, suponiendo que no haya tenido lugar ningún desplazamiento espacial (ya que aún no se ha medido el desplazamiento espacial). Aunque la hipótesis de PVR pueda determinarse empíricamente, una hipótesis de PVR especificada por un usuario que excluya la sobreexploración constituye una buena elección. En la mayor parte de los casos esto impedirá que se utilice vídeo no válido en el algoritmo de registro espacial. Para el vídeo NTSC de 525 líneas muestreado con arreglo a la Recomendación UIT-R BT.601, la sobreexploración ocupa aproximadamente 18 líneas de cuadro en la parte superior e inferior del cuadro, y 22 píxeles a la izquierda y derecha del cuadro. En el vídeo PAL de 625 líneas muestreado con arreglo a la Recomendación UIT-R BT.601, la sobreexploración ocupa aproximadamente 14 líneas de cuadro en la parte superior e inferior del cuadro, y 22 píxeles a la izquierda y a la derecha del cuadro. Cuando se utilicen otros tamaños de imagen (por ejemplo, CIF), debe seleccionarse para ellos una PVR por defecto razonable.

### **6.1.4 Subalgoritmos utilizados por el algoritmo de registro espacial**

El algoritmo de registro espacial utiliza varios subalgoritmos, entre ellos la estimación de la ganancia y desplazamiento de nivel, y la fórmula utilizada para determinar el campo original que más se ajusta a un determinado campo procesado. Estos subalgoritmos se han diseñado para tener un gran rendimiento computacional, ya que el algoritmo de búsqueda iterativo debe ejecutarlos muchas veces.

#### **6.1.4.1 ROI utilizada en todos los cálculos**

El algoritmo efectuará todas las comparaciones de campos entre versiones de una ROI extraída del vídeo procesado, desplazadas espacialmente (para compensar los desplazamientos espaciales introducidos por el HRC) y la correspondiente ROI extraída del vídeo original. La ROI del vídeo procesado desplazada espacialmente se representará por PROI (o sea ROI procesada) y la ROI

correspondiente del vídeo original se representará por OROI (ROI original). Las coordenadas rectangulares que especifican la OROI son fijas a lo largo del algoritmo y se escogen de modo que proporcionen la OROI más grande posible que satisfaga los dos requisitos siguientes:

- La OROI debe corresponder a una PROI que quede dentro de la PVR para todos los posibles desplazamientos espaciales que se examinen.
- La OROI quedará centrada en la imagen original.

#### 6.1.4.2 Ganancia y desplazamiento de nivel

El algoritmo siguiente se utiliza para estimar la ganancia del vídeo procesado. El desplazamiento del campo procesado que se examina se corrige utilizando la estimación actual del desplazamiento espacial. Tras esta corrección de desplazamiento, se selecciona una PROI que corresponda a la OROI fija determinada en el § 6.1.4.1. A continuación, se calculan la desviación típica de los píxeles de luminancia ( $Y$ ) de esta PROI y la desviación típica de los píxeles de luminancia ( $Y$ ) de la OROI. A continuación se estima la ganancia como cociente de la desviación típica de los píxeles de la PROI dividida por la desviación típica de los píxeles de la OROI.

La fiabilidad de esta estimación de ganancia mejora conforme el algoritmo efectúa sus iteraciones aproximándose a los desplazamientos espacial y temporal correctos. Puede utilizarse una ganancia de 1,0 (es decir sin corrección de ganancia) durante los primeros ciclos de iteración. El anterior cálculo de la ganancia es sensible a las degradaciones del vídeo procesado tales como el difuminado. No obstante, a los efectos de registro espacial, esta estimación de ganancia es adecuada porque hace que el vídeo procesado se parezca tanto al vídeo original como sea posible. Para eliminar ganancia del campo procesado, cada píxel de luminancia del campo procesado se divide por la ganancia.

No hay necesidad de determinar ni de corregir el desplazamiento de nivel, ya que el criterio de búsqueda del algoritmo de registro espacial no se ve afectado por los desplazamientos de nivel (véase el § 6.1.4.3).

#### 6.1.4.3 Fórmulas utilizadas para comparar la PROI con la OROI

Tras corregir la ganancia<sup>3</sup> de la PROI (véase el § 6.1.4.2), se utiliza la desviación típica de la imagen diferencia (OROI-PROI) para escoger entre dos o más desplazamientos espaciales o temporales. Se utiliza la estimación de ganancia del mejor emparejamiento anterior para corregir la ganancia de la PROI. Para buscar entre varios desplazamientos espaciales (manteniendo constante el desplazamiento temporal), se calcula la desviación típica de la imagen diferencia (OROI-PROI) para varias PROI generadas con distintos desplazamientos espaciales. Para un determinado campo procesado, se escoge como mejor emparejamiento la combinación de desplazamiento espacial y temporal que produce la menor desviación típica (es decir la mayor cancelación con el original).

#### 6.1.5 Registro espacial con escenas arbitrarias

El registro espacial de un campo procesado de una escena debe examinar una pluralidad de campos originales y de desplazamientos espaciales ya que tanto el campo temporal (es decir el retardo de vídeo) como el desplazamiento espacial son desconocidos. Esto supone que el algoritmo de búsqueda sea complejo y computacionalmente intenso. Además, como el contenido de la escena es

---

<sup>3</sup> A veces es posible omitir la compensación de ganancia para reducir la complejidad computacional. No obstante, sólo se recomienda la omisión de la corrección de ganancia durante las primeras etapas del algoritmo de búsqueda iterativo, cuando el objetivo es encontrar el registro espacial aproximado (véanse, por ejemplo, los § 6.1.5.2 y 6.1.5.3).

arbitrario el algoritmo puede encontrar un registro espacial incorrecto (véase el § 6.1.1). Por consiguiente, lo prudente es calcular el registro espacial de varios campos procesados procedentes de varias escenas diferentes que hayan pasado por el mismo HRC, y combinar los resultados en una estimación válida del desplazamiento espacial. Un solo HRC debe tener un registro espacial constante. De lo contrario, estos desplazamientos espaciales variables en el tiempo se percibirían como degradación (por ejemplo, el vídeo oscilaría de arriba a abajo o de lado a lado). Este apartado describe el algoritmo de registro espacial de abajo a arriba, describiéndose en primer lugar los componentes centrales del algoritmo, y a continuación su aplicación a las escenas de registro espacial y a los HRC.

#### 6.1.5.1 Mejor emparejamiento del campo original en el tiempo

Cuando se realiza el registro espacial utilizando contenidos de escenas, el algoritmo debe encontrar el campo original que más se ajuste al campo actualmente procesado. Por desgracia, puede que el campo original no exista realmente. Por ejemplo, un campo procesado puede contener partes de dos campos originales distintos, ya que puede haber sido interpolado a partir de otros campos procesados. La estimación actual del mejor emparejamiento del campo original (es decir el campo original que más se ajusta al campo actualmente procesado) se mantiene en todas las etapas del algoritmo de búsqueda.

Se adopta la hipótesis inicial de que el primer campo del fichero *YUV* grande procesado se alinea con el primer campo del fichero *YUV* grande original con más menos determinada incertidumbre temporal en los cuadros (representada aquí por *U*). Para cada campo procesado examinado por el algoritmo debe haber una memoria intermedia de *U* cuadros originales antes de este campo y después del mismo. Por consiguiente, el algoritmo comienza examinando los campos procesados que están *U* cuadros dentro del fichero, y examina a continuación uno de cada *F* (frecuencia) campos, deteniéndose *U* cuadros antes del final del fichero.

Los resultados de la última búsqueda del campo procesado anterior (ganancia, desplazamiento vertical y horizontal y desplazamiento temporal) se utilizan para inicializar la búsqueda del campo procesado actual. El cálculo del mejor emparejamiento del campo original con el campo procesado actual se realiza suponiendo un retardo de vídeo constante. Por ejemplo, si se ha determinado que el campo procesado *N* se alinea mejor con el campo original *M* en los ficheros *YUV* grandes, se supondrá que la alineación del campo procesado *N + F* con el campo original *M + F* será mejor al comienzo de la búsqueda.

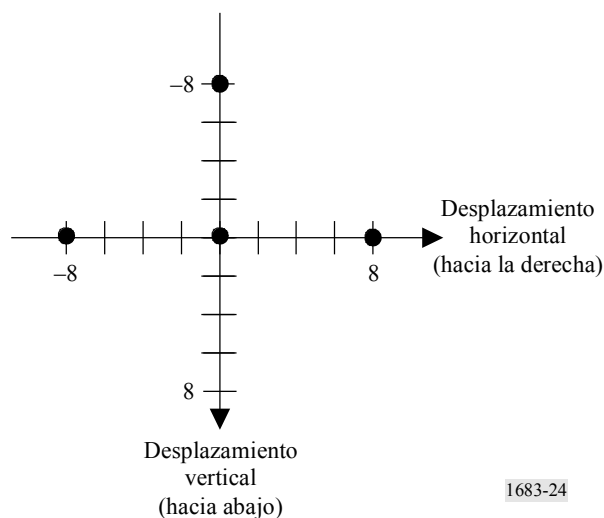
#### 6.1.5.2 Búsqueda general del desplazamiento temporal

La búsqueda exhaustiva de todos los desplazamientos espaciales posibles en toda la incertidumbre temporal para cada campo procesado requeriría un gran número de cálculos. Por este motivo se utiliza una búsqueda multietapa, en la que el primer paso consiste en efectuar una búsqueda general en un conjunto muy limitado de desplazamientos espaciales, a fin de acercarse al campo original emparejado correcto.

En esta búsqueda general se examina el campo uno del cuadro procesado seleccionado (véase la Fig. 23), considerándose únicamente los campos originales de tipo uno separados dos cuadros (es decir separados cuatro campos) en todo el intervalo de más menos la incertidumbre del registro temporal. La búsqueda general considera los cuatro desplazamientos espaciales del vídeo procesado siguientes: sin desplazamiento, ocho píxeles a la izquierda, ocho píxeles a la derecha, y ocho líneas de campo hacia arriba (véase la Fig. 24). En la Fig. 24, los desplazamientos positivos significan que el vídeo procesado está desplazado hacia abajo y a la derecha con respecto al vídeo original. El

desplazamiento de «ocho líneas de campo hacia abajo» no se considera porque las observaciones empíricas han puesto de manifiesto que hay pocos sistemas de vídeo que desplacen la imagen de vídeo hacia abajo. La mejor estimación anterior del desplazamiento espacial (es decir de un campo procesado anteriormente) se incluye también como quinto desplazamiento posible cuando está disponible. El campo original que más se ajusta al campo procesado seleccionado se determina por medio de la técnica de comparación descrita en el § 6.1.4.3. El desplazamiento temporal determinado por el campo original que más se ajusta pasa a ser el punto de partida en la siguiente etapa del algoritmo, búsqueda general del desplazamiento espacial (véase el § 6.1.5.3). De acuerdo con el sistema de coordenadas de la Fig. 21, un desplazamiento temporal positivo significa que el vídeo procesado se ha desplazado en el sentido positivo del tiempo (es decir que el vídeo procesado se ha retrasado con respecto al vídeo original). Con respecto a los ficheros *YUV* grandes original y procesado, un desplazamiento de campo positivo significa por tanto que los campos deben descartarse desde el comienzo del fichero *YUV* grande procesado, mientras que un desplazamiento de campo negativo significa que los campos deben descartarse desde el comienzo del fichero *YUV* grande original.

FIGURA 24  
Desplazamientos espaciales considerados en la búsqueda general del desplazamiento temporal



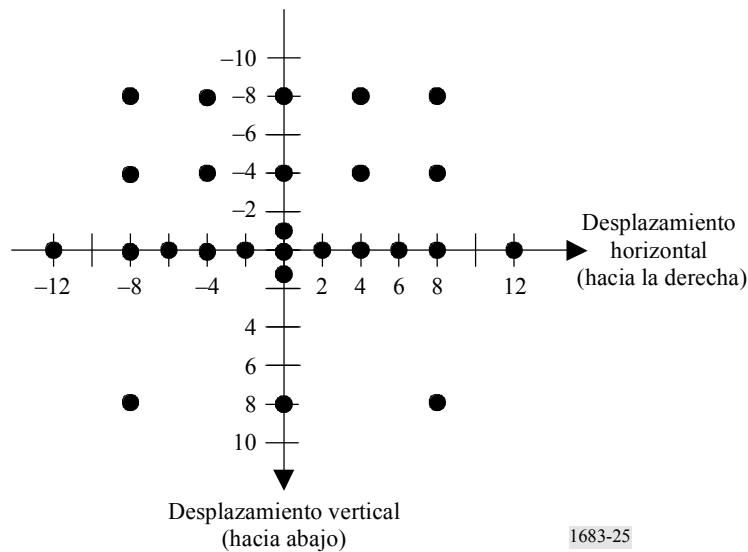
### 6.1.5.3 Búsqueda general del desplazamiento espacial

Con el registro temporal encontrado en la búsqueda general del desplazamiento temporal (véase el § 6.1.5.2), se efectúa ahora una búsqueda general del desplazamiento espacial correcto con un rango más limitado de campos originales. El rango de campos originales considerado en esta búsqueda incluye el campo original tipo uno que más se ajusta (del § 6.1.5.2) y los cuatro campos originales siguientes más aproximados que sean también de tipo uno (campos tipo uno de los dos cuadros anteriores y posteriores al campo original que más se ajusta). La búsqueda general del desplazamiento espacial se efectúa sobre el intervalo de desplazamiento espacial que se muestra en la Fig. 25. Obsérvese que se consideran menos desplazamientos hacia abajo (como en el § 6.1.5.2) debido a que en la práctica son menos probables. El conjunto de desplazamientos espaciales y de campos originales se busca mediante la técnica de comparación descrita en el § 6.1.4.3. Los mejores desplazamientos temporal y espacial resultantes pasan a ser las estimaciones mejoradas para el siguiente paso del algoritmo del § 6.1.5.4.



FIGURA 25

Desplazamientos espaciales considerados en la búsqueda general del desplazamiento espacial



#### 6.1.5.4 Búsqueda detallada del desplazamiento espacio-temporal

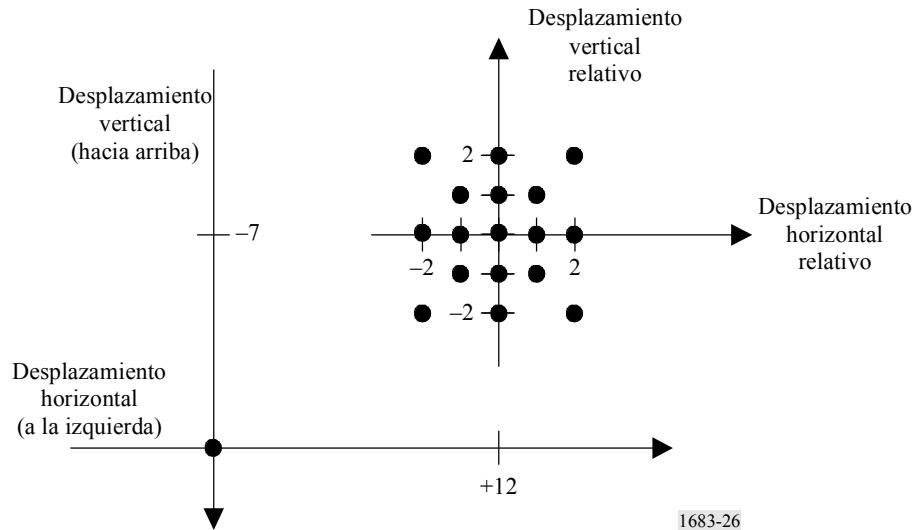
La búsqueda detallada se efectúa sobre un conjunto mucho menor de desplazamientos, centrados en torno a la estimación actual del registro espacial y sólo cinco campos centrados en torno al actual campo original de mejor emparejamiento. Por consiguiente, si el campo original de mejor emparejamiento fuera de tipo uno, la búsqueda se efectuaría sobre tres campos de tipo uno y los dos campos de tipo dos. Entre los desplazamientos espaciales considerados se encuentran la estimación actual del desplazamiento, los ocho desplazamientos que están a menos de un píxel o de una línea de la estimación actual, ocho desplazamientos que están a dos píxeles o a dos líneas de la estimación actual de desplazamiento, y la condición de desplazamiento cero (véase la Fig. 26). En el ejemplo de la Fig. 26, la estimación actual del desplazamiento espacial para el vídeo procesado es de 7 líneas de campo hacia arriba y 12 píxeles a la derecha del vídeo original. El conjunto de desplazamientos espaciales representado en la Fig. 26 forma una búsqueda local casi exhaustiva de los registros espaciales próximos a la estimación de registro espacial. La condición de desplazamiento cero se incluye como verificación de seguridad que contribuye a evitar que el algoritmo se vuelva errático y converja en un mínimo local. El conjunto de desplazamientos espaciales y de campos originales se busca exhaustivamente mediante la técnica de comparación descrita en el § 6.1.4.3. Los mejores desplazamientos espaciales y temporales resultantes pasan a ser las estimaciones mejoradas para el siguiente paso del algoritmo del § 6.1.5.5.

#### 6.1.5.5 Búsquedas detalladas repetitivas

La iteración de la búsqueda detallada del § 6.1.5.4 acercará algo más la estimación actual del desplazamiento espacial ya sea al desplazamiento espacial real o (más difícilmente) a un mínimo falso. Análogamente, una iteración de la búsqueda detallada acercará la estimación actual del campo original mejor alineado ya sea al campo original realmente mejor alineado o (más difícilmente) a un mínimo falso. Por consiguiente, cada búsqueda detallada acercará estas estimaciones a un valor estable. Como las búsquedas detalladas examinan espacial y temporalmente una zona muy limitada, deben efectuarse repetidamente para garantizar que se alcanza la convergencia. Cuando se utilice la compensación de ganancia, la ganancia del campo procesado se estimará de nuevo entre cada dos búsquedas detalladas (véase el § 6.1.4.2).

FIGURA 26

**Desplazamientos espaciales considerados en la búsqueda detallada del desplazamiento espacial**



Las búsquedas detalladas repetitivas se efectúan sobre el campo procesado (véase el § 6.1.5.4) hasta que el mejor desplazamiento espacial y el campo original asociado a dicho desplazamiento espacial permanezcan invariables en dos búsquedas consecutivas. Las búsquedas detalladas repetitivas cesan cuando el algoritmo alterna entre dos desplazamientos espaciales (por ejemplo, un desplazamiento horizontal de 3 y un desplazamiento horizontal de 4, permaneciendo constante lo demás). Esta alternancia se produce cuando la mejor estimación actual del desplazamiento espacial y el campo original asociado a dicho desplazamiento espacial, son idénticos a los de la segunda iteración anterior.

A veces no se produce la convergencia del algoritmo de búsqueda repetitiva. Si el algoritmo no puede converger antes de un determinado número máximo de iteraciones requerido, se termina el algoritmo de búsqueda iterativo y se comunica la condición «búsqueda de desplazamiento fallida» al campo procesado. Este caso especial no suele suponer un problema ya que para cada escena se examinan varios campos procesados (véase el § 6.1.5.6) y para cada HRC se examinan varias escenas (véase el § 6.1.5.7).

### 6.1.5.6 Algoritmo para una escena

Se calcula una estimación inicial de base (es decir de partida) del desplazamiento vertical, del desplazamiento horizontal, y del registro temporal, sin compensación de ganancia alguna del siguiente modo. Se saltan los primeros cuadros procesados de incertidumbre temporal,  $U$ , del fichero  $YUV$  grande. Se efectúa una búsqueda general del desplazamiento temporal sobre el siguiente campo tipo uno procesado (véase el § 6.1.5.2). Obsérvese que esta búsqueda general se efectúa sobre los primeros  $U \cdot 2 + 1$  cuadros de la secuencia de vídeo original para encontrar el campo tipo uno que produzca la mejor alineación. A continuación, se efectúa una búsqueda general del desplazamiento espacial centrada en este campo original de mejor alineación (véase el § 6.1.5.3). A continuación se efectúan hasta cinco búsquedas detalladas espacio-temporales para afinar las estimaciones espaciales y temporales (véanse los § 6.1.5.4 y 6.1.5.5). Si la repetición de estas búsquedas detalladas no consigue encontrar un resultado estable, se descarta este campo procesado del estudio. Se repite el anterior procedimiento para uno de cada  $F$  (frecuencia) cuadros hasta encontrar un campo original de tipo uno que produzca resultados estables. La estimación básica se actualiza periódicamente como se describe a continuación.

Las estimaciones del desplazamiento espacial se calculan para ambos tipos de campo de un cuadro del fichero procesado *YUV* grande del siguiente modo. Utilizando la estimación de base como punto de partida, se ejecutarán tres búsquedas detalladas repetitivas sobre el primer campo procesado de tipo uno. Si la estimación básica es correcta o prácticamente correcta, las búsquedas detalladas repetitivas producirán un resultado estable. En tal caso, el desplazamiento espacial y el retardo temporal correspondientes a dicho campo procesado se almacenan en una matriz destinada a almacenar los resultados del campo uno. Si no se encuentra un resultado estable, lo más probable es que el desplazamiento espacial sea correcto aunque la estimación del desplazamiento temporal esté desplazada (o sea, la estimación actual del desplazamiento temporal está a más de dos cuadros de distancia del desplazamiento temporal verdadero). Así pues se realiza una búsqueda general del desplazamiento temporal que incluya la mejor estimación actual del desplazamiento espacial. Esta búsqueda general corregirá normalmente la estimación de retardo temporal. Cuando se completa la búsqueda general del desplazamiento temporal, su salida se utiliza como punto de partida, y se efectúan hasta cinco búsquedas detalladas repetidas. Si esta segunda búsqueda repetida no consigue encontrar un resultado estable, se comunicará el fallo del registro espacial para este cuadro (es decir tanto para el campo tipo uno como para el campo tipo dos). Si se encuentra un resultado estable en esta segunda búsqueda, el desplazamiento espacial y el retardo temporal para este campo se almacenan en la matriz del campo uno. Asimismo, se actualizan el desplazamiento espacial y el retardo temporal utilizados como punto de partida para el siguiente campo procesado de tipo uno (es decir para el primer campo procesado, se utilizan los resultados de base y después se utiliza el último resultado estable). Una vez estimado el desplazamiento espacial para el primer campo procesado de tipo uno, se estima el desplazamiento espacial para el primer campo procesado de tipo dos. Utilizando los resultados espaciales del campo uno como punto de partida, se utilizan los mismos pasos para encontrar el desplazamiento espacial del campo dos (o sea tres búsquedas detalladas y si es necesario una búsqueda general del desplazamiento temporal seguida de cinco búsquedas detalladas repetidas). Si se encuentra un resultado estable para el campo dos, se almacenan el desplazamiento vertical y el horizontal del campo dos en una matriz diferente destinada a almacenar los resultados del campo dos.

El procedimiento descrito en el párrafo anterior se aplica al cálculo del desplazamiento espacial de los dos tipos de campo de uno de cada  $F$  (frecuencia) cuadros del fichero *YUV* grande que contiene el vídeo procesado. Los primeros cuadros procesados de incertidumbre temporal,  $U$ , del fichero *YUV* grande se saltan. Esta secuencia de cálculos se utiliza entonces para calcular estimaciones válidas del desplazamiento espacial para cada tipo de campo de la escena objeto de examen. Se ordenan los resultados del campo uno correspondientes a cada cuadro, y se retiene el percentil 50 como desplazamiento vertical global del campo uno. Análogamente, se ordenan los resultados de desplazamiento vertical del campo dos y se toma el percentil 50 como desplazamiento vertical global del campo dos. Se ordenan los resultados del desplazamiento horizontal del campo uno de cada cuadro, y se toma el percentil 50 como desplazamiento horizontal global. Las posibles diferencias entre el desplazamiento horizontal del campo uno y el del campo dos se deben probablemente al desplazamiento horizontal de los subpíxeles (por ejemplo, a un desplazamiento horizontal de 0,5 píxeles). Los desplazamientos horizontales de los subpíxeles producirán estimaciones que incluyan los dos desplazamientos más próximos. Utilizando el percentil 50 se puede elegir el desplazamiento horizontal más probable, lo que produce una precisión de registro espacial que es válida hasta los 0,5 píxeles más próximos<sup>4</sup>.

---

<sup>4</sup> El registro espacial hasta los 0,5 píxeles más próximos es suficiente para las mediciones de la calidad de vídeo descritas en este Anexo. Las técnicas de registro espacial de subpíxeles van más allá del propósito del presente Anexo.

### 6.1.5.7 Algoritmo para un HRC

Si se pasan varias escenas por el mismo HRC, los resultados del registro espacial de las escenas deben ser idénticos. Por consiguiente, el filtrado de los resultados obtenidos de varias escenas puede mejorar la validez y precisión de las mediciones del desplazamiento espacial. Los resultados globales del registro espacial del HRC pueden utilizarse entonces para compensar todo el vídeo procesado en dicho HRC.

### 6.1.5.8 Observaciones sobre el algoritmo

Algunas escenas de vídeo no son adecuadas para la estimación del registro espacial. Algunas veces, el algoritmo descrito localizará un mínimo falso. Otras, el algoritmo pasará de una solución a otra sin llegar a alcanzar un resultado estable. Por ello es aconsejable examinar varias imágenes de la misma escena y aplicar un filtro de mediana a estos resultados en distintas escenas (o sea ordenar los resultados de menor a mayor y seleccionar el percentil 50). El registro espacial por el algoritmo de escena constituye un algoritmo heurístico que utiliza patrones de desplazamiento espacial que se hayan observado a partir de un muestreo de sistemas de vídeo. Estas hipótesis pueden llegar a ser incorrectas en determinados sistemas, haciendo que los algoritmos encuentren un desplazamiento espacial incorrecto. Hay que tener en cuenta, no obstante, que los fallos del algoritmo tienden a producir desplazamientos espaciales incoherentes de un cuadro a otro y de una escena a otra (es decir cuando falla el algoritmo suele producir resultados dispersos). Cuando el algoritmo produce los mismos desplazamientos espaciales o desplazamientos espaciales muy semejantes, para cada escena, el grado de confianza es elevado. Cuando los resultados de un campo individual de una escena son erráticos, el grado de confianza es bajo.

### 6.1.6 Registro espacial del vídeo progresivo

El registro espacial del vídeo progresivo sigue los mismos pasos que los algoritmos entrelazados, con modificaciones de menor importancia. Cuando los algoritmos entrelazados operan sobre el campo uno y el campo dos por separado, el algoritmo progresivo opera sobre cuadros. Por consiguiente, se ignoran todas las referencias al campo dos y, exceptuando las búsquedas detalladas, se duplica el rango de desplazamiento vertical.

La modificación del rango de desplazamiento vertical es de la mayor importancia para el desplazamiento espacial amplio. Cuando se efectúa una búsqueda general del desplazamiento espacial (véase el § 6.1.5.3) los números del eje vertical de la Fig. 25 deben duplicarse (o sea, +8 pasa a ser +16 y -4 pasa a ser -8)<sup>5</sup>. Por otra parte, para las imágenes CIF y QCIF progresivas, los rangos de búsqueda espacial general horizontal y vertical se dividen por dos debido a los menores desplazamientos que se suelen encontrar con estos tamaños de imagen. Por ejemplo, utilizando imágenes CIF en la Fig. 25, el eje horizontal se extendería de -6 a +6 píxeles y el vertical de -8 a +8 líneas de cuadro.

El rango de búsqueda temporal, expresado en cuadros, permanece prácticamente inalterado. En la búsqueda temporal general del § 6.1.5.2, en vez de buscar la concordancia de un campo uno procesado con uno de cada dos campos uno originales, el algoritmo progresivo compara un cuadro procesado con uno de cada dos cuadros originales. En el algoritmo de barras de color, la búsqueda examina los desplazamientos espaciales de un cuadro procesado y de un cuadro original (es decir no hay búsqueda temporal).

El único paso que requiere modificaciones más complejas es la búsqueda detallada del § 6.1.5.4. En ésta, los desplazamientos verticales quedan inalterados, y oscilan entre -2 líneas de cuadro y

---

<sup>5</sup> En una posible excepción a esta duplicación, el desplazamiento espacial asociado a cero píxeles horizontalmente y más menos una línea de campo verticalmente podría quedar como más menos una línea de campo verticalmente. Son comunes los desplazamientos espaciales muy próximos a (cero, cero).

+2 líneas de cuadro. Por consiguiente, el eje vertical de la Fig. 26 se interpreta como referido a líneas de cuadro. La amplitud temporal de esta búsqueda detallada puede establecerse en cinco cuadros originales centrados en el actual cuadro original alineado, en vez de los tres cuadros originales del otro sistema. La ampliación de la búsqueda a cinco cuadros puede mejorar la velocidad y el rendimiento de la búsqueda detallada en comparación con la versión entrelazada del algoritmo, ya que es más probable que los HRC progresivos tengan retardo de vídeo variable que desplazamientos espaciales no nulos.

Teniendo en cuenta las modificaciones algorítmicas para los sistemas de vídeo progresivo, muchos de los parámetros de búsqueda de desplazamiento espacial pueden modificarse sin que ello afecte a la integridad del algoritmo. Por ejemplo, pueden considerarse desplazamientos espaciales distintos de cero píxeles y de cero líneas para la búsqueda temporal general. El desplazamiento espacial a cero píxeles horizontalmente y a 8 líneas de campo verticalmente para vídeo entrelazado podría trasladarse a 16 líneas de cuadro para vídeo progresivo, como se ha recomendado anteriormente, o colocarse a 8 líneas de cuadro, en la hipótesis de que resulte improbable que las secuencias de vídeo contengan 16 líneas de cuadro de desplazamiento vertical. Análogamente, el desplazamiento espacial a cero líneas verticalmente y 8 píxeles horizontalmente podría trasladarse a 9 ó 10 píxeles horizontalmente sin efecto pernicioso alguno. En otro ejemplo, el número exacto de búsquedas detalladas consecutivas a efectuar podría aumentarse o disminuirse para aplicaciones específicas. La exactitud de los valores aquí recomendados es bastante menos importante que la estructura real del algoritmo de búsqueda.

## 6.2 Región válida

El vídeo NTSC (de 525 líneas) y el PAL (de 625 líneas) muestreados con arreglo a la Recomendación UIT-R BT.601 pueden tener un borde de píxeles y líneas que no contenga una imagen válida. El vídeo original de la cámara puede llenar solamente una parte del cuadro de la Recomendación UIT-R BT.601. Un sistema de vídeo digital que utilice compresión puede reducir aún más la zona de imagen a fin de ahorrar bits de transmisión. Si los píxeles y líneas no transmitidos corresponden a la zona de sobreexploración de la imagen de televisión, el usuario final típico no debe percibir la omisión de líneas y píxeles. Si estos píxeles y líneas no transmitidos desbordan el área de sobreexploración, el espectador puede percibir un borde negro en torno a la imagen, ya que el sistema insertará normalmente negro en esta zona de imagen no transmitida. Los sistemas de vídeo (especialmente los que ejecutan un filtrado paso bajo) pueden presentar una gradación ascendente desde el borde negro hasta la zona de imagen. Estos efectos de transición suelen presentarse más frecuentemente en los lados izquierdo y derecho de la imagen aunque también pueden aparecer en su parte superior e inferior. Ocasionalmente, el vídeo procesado puede contener asimismo varias líneas de vídeo corrompido en la parte superior o inferior de la imagen que pueden ser invisibles para el espectador (por ejemplo las grabadoras de vídeo de VHS corrompen varias líneas del fondo de la imagen en la zona de sobreexploración). Para evitar que las zonas exteriores a la imagen influyan en las mediciones de la VQM, estas zonas deben excluirse de la medición de la VQM. El algoritmo automático de región válida aquí presentado estima la región válida de las secuencias de vídeo original y procesada de modo que los cálculos subsiguientes no consideren las líneas corrompidas en la parte superior e inferior del cuadro de la Recomendación UIT-R BT.601, los píxeles de los bordes negros ni los efectos de transición cuando los bordes negros alcanzan la zona de imagen.

### 6.2.1 Algoritmo central de la región válida

En este punto se describe el algoritmo central de la región válida aplicable a una única imagen original o procesada. Este algoritmo requiere tres argumentos de entrada: una imagen, una región válida máxima, y la estimación de la región válida actual.

- *Imagen:* el algoritmo central utiliza la imagen de luminancia de la Recomendación UIT-R BT.601 correspondiente a un único cuadro de vídeo. Cuando se mide la región válida de una secuencia vídeo procesado, cualquier desplazamiento espacial impuesto por el sistema de vídeo debe haber sido suprimido de la imagen de luminancia antes de aplicar el algoritmo central (véase el § 6.1).
- *Región válida máxima:* el algoritmo central no considerará los píxeles y las líneas exteriores a la región de vídeo válida máxima. Esto constituye un mecanismo para que el usuario especifique una región válida máxima que sea inferior a la totalidad de la zona de imagen si la información a priori indica que la imagen muestreada tiene píxeles o líneas corrompidos como se ha expuesto en el § 6.2.
- *Región válida actual:* la región válida actual es una estimación de la región válida y queda totalmente dentro de la región válida máxima. Se sabe que todos los píxeles que se encuentran dentro de la región válida actual contienen vídeo válido; los píxeles que se encuentran fuera de la región válida actual pueden tener contenidos de vídeo válidos o no. Inicialmente, la región válida actual se establece como la zona más pequeña posible situada en el centro exacto de la imagen.

El algoritmo central examina el área de vídeo comprendida entre la región válida máxima y la región válida actual. Si alguno de estos píxeles parece contener vídeo válido, se aumenta la estimación de la región válida actual. A continuación, se describe el algoritmo en detalle para el borde izquierdo de la imagen.

*Paso 1:* Se calcula el promedio de la columna extremo izquierda de píxeles de la región válida máxima. La columna extrema izquierda de píxeles se representará por columna  $J-1$  y el promedio se representará por  $M_{J-1}$ .

*Paso 2:* Se toma la media de la siguiente columna de píxeles,  $M_J$ .

*Paso 3:* La columna  $J$  se declara vídeo no válido si es negra, ( $M_J < 20$ ) o si el nivel medio de píxeles del valor medio de las columnas sucesivas indica una gradación ascendente desde el borde negro hasta la imagen válida ( $M_J - 2 > M_{J-1}$ ). De ser cierta alguna de estas condiciones, se aumenta  $J$  y se repiten los Pasos 2 y 3. De lo contrario, se va al Paso 4.

*Paso 4:* Si la última columna  $J$  queda dentro de la región válida, es que no se ha obtenido ninguna información nueva. De lo contrario, se actualiza la región válida actual con  $J$  como coordenada izquierda.

El algoritmo para encontrar el borde superior de la imagen es semejante al aquí expuesto para el borde izquierdo. Para los bordes inferior y derecho,  $J$  se decrementa en vez de incrementarse; por lo demás el algoritmo es idéntico. Los valores obtenidos para la parte superior, la izquierda, la inferior, y la derecha indican el último píxel o línea válidos.

Las condiciones de detención definidas en el Paso 3 pueden resultar inútiles en función del contenido de la escena. Por ejemplo, una imagen que contenga un negro auténtico en el lado izquierdo (es decir un negro que forme parte de la escena) hará que el algoritmo central deduzca que la columna de vídeo válida más a la izquierda esté más cerca del centro de la imagen de lo que debería. Por esta razón, el algoritmo central se aplica a varias imágenes de una secuencia de vídeo, incrementando de este modo la precisión de la estimación de la región válida.

## 6.2.2 Aplicación del algoritmo central de región válida a una secuencia de vídeo

### 6.2.2.1 Vídeo original

El algoritmo central se aplica en primer lugar a la secuencia original de imágenes. Para vídeo NTSC muestreado con arreglo a la Recomendación UIT-R BT.601 (véase el § 5), los valores recomendados para la región válida máxima son superior = 6, izquierda = 6, inferior = 482, derecha = 714.

Para vídeo PAL muestreado con arreglo a la Recomendación UIT-R BT.601, los valores recomendados para la región válida máxima son superior = 6, izquierda = 16, inferior = 570, derecha = 704. El algoritmo central se ejecuta sobre la primera imagen de la secuencia de vídeo y después sobre una de cada  $F$  (frecuencia) imágenes. Por ejemplo, si la frecuencia especificada es 15, el algoritmo central examinará los números de secuencia de imágenes 0, 15, 30, 45 y así sucesivamente. Una vez examinadas todas las imágenes de la secuencia, la región válida actual contendrá la zona válida más grande obtenida de todas las imágenes examinadas de la secuencia de vídeo. Se considera que los píxeles y las líneas comprendidos entre esta última región válida actual y la región válida máxima contienen o bien negro o bien una gradación ascendente desde el negro.

La última región válida debe contener un número par de líneas y un número par de píxeles. Las eventuales coordenadas impares superior o izquierda se incrementan en uno. Por consiguiente, si la región tiene un número impar de líneas, se disminuye la parte inferior; análogamente, si la región tiene un número impar de píxeles (es decir horizontalmente), se decrementa la derecha. Esto simplifica el procesamiento de color del vídeo muestreado con arreglo a la Recomendación UIT-R BT.601, ya que los canales de color se submuestran por 2 cuando se comparan con el canal de luminancia. Además, cada campo entrelazado de vídeo contendrá el mismo número de líneas de vídeo. Esto garantiza que las subregiones espacial y temporal (de las que se extraerán las características) contengan siempre vídeo válido con iguales contribuciones de ambos campos entrelazados. La región válida resultante se devuelve como región válida original.

#### 6.2.2.2 Vídeo procesado

Cuando se calcula la región válida de la secuencia de vídeo procesada, los valores de la región válida máxima del algoritmo central se hacen iguales, en primer lugar, a la región válida original correspondiente encontrada para esta escena. A continuación se reduce el tamaño de la región válida máxima en los píxeles y líneas que se consideren no válidos debido a la corrección del desplazamiento espacial de los cuadros de vídeo procesados. A continuación se ejecuta el algoritmo central sobre la primera imagen de la secuencia de vídeo procesada y sobre una de cada  $F$  (frecuencia) imágenes subsiguientes (o sea si frecuencia =  $F$ , se utilizan las imágenes  $Y(0)$ ,  $Y(F)$ ,  $Y(2F)$ ,  $Y(3F)$ , y así sucesivamente).

Una vez aplicado el algoritmo central a la secuencia de vídeo procesada, la región válida encontrada por el algoritmo central se reduce hacia el interior en un margen de seguridad. El margen de seguridad recomendado consiste en descartar una línea de la parte superior y del fondo y cinco píxeles de la izquierda y de la derecha. Los grandes márgenes de seguridad a la izquierda y a la derecha garantizan que cualquier gradación ascendente o descendente desde el negro queda excluida de la región válida procesada.

La región válida procesada definitiva debe contener un número par de líneas y un número par de píxeles. Las eventuales coordenadas impares superior o inferior se incrementan en uno. Por consiguiente, si la región contiene un número impar de líneas, se decrementa la parte inferior; análogamente si la región contiene un número impar de píxeles (es decir horizontalmente), se decrementa la derecha. La región válida resultante se devuelve como región válida procesada.

#### 6.2.3 Observaciones sobre el algoritmo de región válida

Este algoritmo automático de región válida funciona bien para estimar la región válida en la mayoría de las escenas. Debido a que las posibilidades de contenidos de las escenas son prácticamente infinitas, el algoritmo aquí descrito adopta una solución conservadora para estimar la región válida. Un examen manual de la región válida produciría probablemente una región mayor. Las estimaciones conservadoras de la región válida son más adecuadas para los sistemas automáticos de medición de la calidad de vídeo, porque el descarte de una pequeña cantidad de vídeo

tiene escasa repercusión sobre la estimación de la calidad y en todo caso este vídeo suele corresponder a la parte de sobreexploración del vídeo. Por otra parte, la inclusión de vídeo corrupto en los cálculos de la calidad de vídeo puede tener una gran repercusión sobre la estimación de la calidad.

Este algoritmo no tiene suficiente inteligencia artificial para distinguir entre los píxeles y líneas corrompidos del borde de una imagen y el contenido auténtico de una escena. Por ello se utiliza una regla práctica consistente en suponer que dicho vídeo no válido suele corresponder generalmente a los bordes extremos de la imagen. La especificación de una región válida máxima de vídeo definida por el usuario (es decir el punto de arranque del algoritmo automático) proporciona un mecanismo de exclusión del estudio de estos bordes de imagen posiblemente corrompidos.

Cuando se aplica el algoritmo de región válida a vídeo no muestreado con arreglo a la Recomendación UIT-R BT.601 (es decir de formato intermedio común, o CIF, utilizado por la Recomendación UIT-T H.261), los valores recomendados para la región válida máxima cuando se examina el vídeo original son los de la imagen completa. En estos casos, el vídeo muestreado no contiene normalmente sobreexploración corrompida, de modo que resulta superflua una región válida máxima menor que la imagen completa.

### 6.3 Ganancia y desplazamiento

#### 6.3.1 Algoritmo central de ganancia y desplazamiento de nivel

Este apartado explica el método de calibración de la ganancia y del desplazamiento de nivel. Es requisito previo a la aplicación de este algoritmo que las imágenes original y procesada estén registradas espacialmente (véase el § 6.1). Las imágenes original y procesada deben estar asimismo registradas temporalmente, lo que se abordará más adelante en el § 6.4. La calibración de la ganancia y del desplazamiento de nivel puede ejecutarse sobre los campos o sobre los cuadros, según proceda.

El método aquí presentado adopta la hipótesis de que las señales  $Y$ ,  $C_B$ , y  $C_R$  de la Recomendación UIT-R BT.601 tienen ganancias y desplazamientos de nivel independientes. Esta hipótesis bastará normalmente para calibrar los sistemas de vídeo de componentes (es decir  $Y$ ,  $R-Y$ ,  $B-Y$ ). No obstante, en los sistemas de vídeo compuesto o  $S$ -vídeo, es posible tener una rotación de fase de la información de crominancia ya que las dos componentes de crominancia están multiplexadas en un vector de señal complejo con amplitud y fase. El algoritmo aquí presentado no calibrará adecuadamente los sistemas de vídeo que introduzcan una rotación de fase de la información de crominancia (o sea, los aparatos de televisión con ajuste de tono).

- Como se ha indicado anteriormente, en este modelo de calibración se supone que no hay acoplamiento cruzado entre ninguna de las tres componentes de vídeo. De acuerdo con esta hipótesis, el algoritmo central de calibración se aplica independientemente a cada uno de los tres canales:  $Y$ ,  $C_B$ , y  $C_R$ .
- La región válida del plano de imagen original y procesada se divide en primer lugar en  $N$  subregiones. Se calculan los valores medios *original* y *procesado* correspondientes a cada una de estas subregiones (es decir media sobre espacio). A continuación, estos valores *original* y *procesado* se representan como vectores columna de  $N$  elementos  $\underline{Q}$  y  $\underline{P}$ , respectivamente:

$$\underline{Q}_{N \times 1} = \begin{bmatrix} \text{original}_1 \\ \cdot \\ \cdot \\ \cdot \\ \text{original}_N \end{bmatrix}, \quad \underline{P}_{N \times 1} = \begin{bmatrix} \text{procesado}_1 \\ \cdot \\ \cdot \\ \cdot \\ \text{procesado}_N \end{bmatrix}$$



La calibración consiste en calcular la ganancia,  $g$ , y el nivel de desplazamiento,  $l$ , con arreglo al siguiente modelo:

$$\underline{P} = g\underline{O} + l$$

Dado que sólo hay dos incógnitas (es decir  $g$  y  $l$ ) pero  $N$  ecuaciones (es decir  $N$  subregiones), hay que resolver un sistema sobredeterminado de ecuaciones lineales dado por:

$$\hat{\underline{P}} = A \begin{bmatrix} l \\ g \end{bmatrix}$$

siendo  $A$  una matriz  $N \times 2$  dada por  $A_{N \times 2} = [\underline{1} \quad \underline{O}]$ , y  $\underline{1}$  un vector columna de  $N$  elementos de «1» dado por:

$$\underline{1}_{N \times 1} = \begin{bmatrix} 1_1 \\ \cdot \\ \cdot \\ \cdot \\ 1_N \end{bmatrix}$$

$\hat{\underline{P}}$  es la estimación de las muestras procesadas si se aplica la corrección de ganancia y desplazamiento de nivel a las muestras originales. La solución de mínimos cuadrados para este problema sobredeterminado (con tal de que  $N > 2$ ) viene dada por:

$$\begin{bmatrix} l \\ g \end{bmatrix} = (A^T A)^{-1} A^T P$$

donde el superíndice, T, indica la matriz transpuesta y el superíndice, -1, indica la matriz inversa.

Cuando se aplica el algoritmo central de ganancia y desplazamiento independientemente a cada uno de los tres canales, se obtienen seis resultados: ganancia  $Y$ , desplazamiento  $Y$ , ganancia  $C_B$ , desplazamiento  $C_B$ , ganancia  $C_R$ , y desplazamiento  $C_R$ .

### 6.3.2 Utilización de escenas

El algoritmo básico del § 6.3.1 puede aplicarse a secuencias de vídeo originales y procesadas con tal de que se hayan registrado espacial y temporalmente. Esta técnica basada en escenas divide la imagen en bloques adyacentes con niveles de intensidad desconocidos. Se recomienda un tamaño de subregión de 16 líneas  $\times$  16 píxeles para los cuadros (es decir 8 líneas  $\times$  16 píxeles para un campo Y NTSC o PAL y 8 líneas  $\times$  8 píxeles para  $C_B$  y  $C_R$  debido al submuestreo de los planos de imagen de color). Se calcula la media en el espacio de las muestras  $[Y, C_B, C_R]$  para cada subregión, o bloque, correspondiente original o procesada para formar una imagen espacialmente submuestreada. Todos los bloques seleccionados deben quedar dentro de la PVR.

#### 6.3.2.1 Registro de las imágenes procesadas

Para mayor simplicidad, se supondrá encontrado el mejor registro espacial gracias a una de las técnicas expuestas en el § 6.1. Antes de la estimación de la ganancia y del desplazamiento de nivel, cada imagen procesada debe registrarse también temporalmente. Para el cálculo de la ganancia y del desplazamiento de nivel debe utilizarse la imagen original que mejor se alinee con la imagen procesada. Si el retardo de vídeo es variable, este registro temporal debe ejecutarse para cada imagen procesada. Si el retardo de vídeo es constante para la escena, el registro temporal sólo necesita ejecutarse una vez.

Para registrar temporalmente una imagen procesada, se crearán en primer lugar los campos (o cuadros en el vídeo progresivo) original y procesado, submuestreados espacialmente especificados en el § 6.3.2, tras corregir el desplazamiento espacial del vídeo procesado. Utilizando las imágenes  $Y$  submuestreadas, se aplicará a la función de búsqueda del § 6.1.4.3, pero efectuando esta búsqueda sobre todas las imágenes originales que queden dentro de la incertidumbre del registro temporal,  $U$ . Se utilizará el mejor registro temporal resultante para los tres planos de imagen,  $Y$ ,  $C_B$ , y  $C_R$ .

### 6.3.2.2 Ganancia y desplazamiento de nivel de las imágenes registradas

Se utiliza una solución iterativa de mínimos cuadrados con una función de costo para contribuir a la minimización de la repercusión de los valores atípicos en el ajuste. Esto se debe a que los valores atípicos proceden normalmente de distorsiones y no de simples variaciones del desplazamiento de nivel y de la ganancia, y la asignación del mismo peso a estos valores atípicos distorsionaría el ajuste.

El siguiente algoritmo se aplica por separado a los  $N$  píxeles concordantes originales y procesados de cada una de las tres imágenes submuestreadas espacialmente [ $Y$ ,  $C_B$ ,  $C_R$ ].

*Paso 1:* Utilizar la solución normal de mínimos cuadrados del § 6.3.1 para generar la estimación inicial del desplazamiento de nivel y ganancia:  $\begin{bmatrix} l \\ g \end{bmatrix} = (A^T A)^{-1} A^T \underline{P}$ .

*Paso 2:* Generar un vector de error,  $\underline{E}$ , igual al valor absoluto de la diferencia entre las muestras procesadas auténticas y las muestras procesadas ajustadas:  $\underline{E} = |\underline{P} - \hat{P}|$ .

*Paso 3:* Generar un vector de costos,  $\underline{C}$ , que sea recíproco, elemento a elemento, del vector de errores,  $E$ , más un pequeño épsilon,  $\epsilon$ :  $\underline{C} = \frac{1}{E + \epsilon}$ . El  $\epsilon$  impide la división por cero y establece el peso relativo de un punto que esté en la línea ajustada al de otro que esté fuera de la línea ajustada. Se recomienda un  $\epsilon$  de 0,1.

*Paso 4:* Normalizar el vector de costos  $C$  para la norma unitaria (es decir cada elemento de  $C$  se divide por la raíz cuadrada de la suma de los cuadrados de todos los elementos de  $C$ ).

*Paso 5:* Generar el vector de costos  $C^2$  que sea el cuadrado, elemento a elemento, del vector de costos  $C$  del Paso 4.

*Paso 6:* Generar una matriz diagonal  $N \times N$  de costos,  $C^2$ , que contenga los elementos del vector de costos,  $C^2$ , dispuesto sobre la diagonal, relleno el resto a ceros.

*Paso 7:* Utilizar la matriz diagonal de costos,  $C^2$ , del Paso 6, y ejecutar el ajuste de mínimos cuadrados ponderado en costo para determinar la siguiente estimación del desplazamiento de nivel y de la ganancia:  $\begin{bmatrix} l \\ g \end{bmatrix} = (A^T C^2 A)^{-1} A^T C^2 \underline{P}$ .

*Paso 8:* Repetir los Pasos 2 a 7 hasta que las estimaciones del desplazamiento de nivel y de la ganancia converjan con cuatro posiciones decimales.

Estos Pasos se aplican por separado al campo procesado uno y al campo procesado dos, para obtener las estimaciones de  $g$  y  $l$ . El campo uno y el dos deben examinarse por separado, debido a que no es necesario que los campos originales registrados temporalmente correspondan a un cuadro de la secuencia de vídeo original. Para el vídeo progresivo, los Pasos anteriores se aplican a todo el cuadro procesado al mismo tiempo.

### 6.3.2.3 Estimación de la ganancia y del desplazamiento de nivel para una secuencia de vídeo y HRC

El algoritmo descrito anteriormente se aplica a varios pares de campos originales y procesados concordantes, distribuidos cada  $F$  (frecuencia) cuadros de la escena (para vídeo progresivo, pares de cuadros originales y procesados). A continuación se aplica un filtro de mediana a las seis historias temporales de los desplazamientos de nivel y ganancias para obtener estimaciones medias para cada escena.

Si se han pasado varias escenas por el mismo HRC, el nivel de desplazamiento y la ganancia de las escenas se considerará idéntico. Así pues, el filtrado de los resultados obtenidos de varias escenas puede mejorar la validez y precisión de las mediciones del desplazamiento de nivel y de la ganancia. Los resultados globales del desplazamiento de nivel y de la ganancia del HRC pueden utilizarse entonces para compensar todo el vídeo procesado en dicho HRC.

### 6.3.3 Aplicación de las correcciones de la ganancia y del desplazamiento de nivel

Los algoritmos de registro temporal (véase el § 6.4) y la mayor parte de las características de calidad (véase el § 7) recomendarán la supresión de la ganancia aquí calculada. Para suprimir la ganancia y el desplazamiento de nivel del plano Y, se aplicará la siguiente fórmula a cada píxel procesado:

$$\text{Nueva } Y(i, j, t) = [Y(i, j, t) - 1] / g$$

La corrección de la ganancia y del nivel de desplazamiento no se ejecuta en los planos de color (es decir  $C_B$  y  $C_R$ ). Los errores de crominancia de percepción se capturan, en cambio, por la métrica del color. Los planos de imagen  $C_B$  y  $C_R$  pueden corregirse en ganancia y desplazamiento de nivel a los efectos de visualización.

## 6.4 Registro temporal

Los modernos sistemas digitales de comunicación de vídeo, suelen requerir varias decenas de segundo para procesar y transmitir el vídeo desde la cámara emisora hasta la pantalla receptora. Los retardos de vídeo excesivos impiden la eficacia de la comunicación bidireccional. Por consiguiente, los métodos objetivos de medición extremo a extremo del retardo de las comunicaciones de vídeo son importantes para los usuarios finales a efectos de la especificación y comparación de los servicios, y para que los proveedores de equipos y servicios optimicen y mantengan su oferta de productos. El retardo de vídeo puede depender de los atributos dinámicos de la escena original (por ejemplo, el detalle espacial, el movimiento) y del sistema de vídeo (por ejemplo de la velocidad binaria). Por citar un ejemplo, las escenas con mucho movimiento pueden sufrir más retardo de vídeo que aquéllas con poco movimiento. Por consiguiente, las mediciones del retardo de vídeo deben realizarse en servicio para poder ser representativas y precisas. Las estimaciones del retardo de vídeo son necesarias para alinear temporalmente las características del vídeo original y procesado representadas en la Fig. 19 antes de realizar las mediciones de calidad.

Algunos sistemas de transmisión de vídeo pueden proporcionar información de sincronización (por ejemplo, pueden etiquetarse los cuadros original y procesado con algún tipo de esquema de numeración de cuadros). En general, no obstante, debe medirse la sincronización entre las secuencias de vídeo original y procesada. Este apartado presenta una técnica de estimación del retardo de vídeo basada en los cuadros de vídeo original y procesado. Esta técnica se puede calificar como «basada en cuadros» por apoyarse en la correlación de imágenes de menor resolución, submuestreadas en el espacio y extraídas de las secuencias de vídeo original y procesado. Esta técnica basada en cuadros estima el retardo de cada cuadro o campo (en los sistemas de vídeo entrelazados). Estas estimaciones individuales se combinan para estimar el retardo medio de la secuencia de vídeo.

#### 6.4.1 Algoritmo basado en tramas para estimar los retardos temporales variables entre las secuencias de vídeo original y procesada

En este punto se describe un algoritmo de registro temporal basado en cuadros. Para reducir la influencia de las distorsiones sobre el registro temporal, las imágenes se submuestran espacialmente y se normalizan de modo que la varianza sea la unidad. Este algoritmo registra temporalmente cada imagen procesada por separado, localizando la imagen original más semejante. Algunas de estas mediciones individuales del registro temporal puede ser incorrecta pero estos errores tenderán a distribuirse aleatoriamente. Cuando se combinan las mediciones de retardo de una serie de imágenes por medio de un esquema de votación, la estimación global del retardo medio de una secuencia de vídeo llega a ser bastante precisa. Este algoritmo de registro temporal no utiliza porciones de escena en reposo o prácticamente sin movimiento, ya que las imágenes originales son aproximadamente idénticas entre sí.

##### 6.4.1.1 Constantes que utiliza el algoritmo

- BELOW\_WARN:** Umbral utilizado para examinar las correlaciones a fin de decidir si el máximo de correlación secundaria es suficientemente grande como para indicar registro temporal ambiguo. Se recomienda un BELOW\_WARN de 0,9.
- BLOCK\_SIZE:** Factor de submuestreo. Se especifica en líneas de cuadro verticalmente y en píxeles horizontalmente. Se recomienda un BLOCK\_SIZE de 16.
- DELTA:** Los máximos secundarios de la curva de correlación que están dentro de DELTA de la máxima (mejor) correlación se ignoran. Se recomienda una DELTA de 4.
- HFw:** Mitad de la anchura del filtro utilizado para suavizar el histograma de los valores de registro temporal cuadro a cuadro. Se recomienda un HFw de 3.
- STILL\_THRESHOLD:** Umbral utilizado para detectar escenas de vídeo en reposo (el registro temporal basado en tramas no puede utilizarse en secciones de vídeo en reposo). Se recomienda un STILL\_THRESHOLD de 0,002.

##### 6.4.1.2 Entradas al algoritmo

Una secuencia de  $N$  imágenes de luminancia de vídeo original:  $Y_O(t)$ ,  $0 \leq t < N$ <sup>6</sup>.

Una secuencia de  $N$  imágenes de luminancia de vídeo procesado:  $Y_P(t)$ ,  $0 \leq t < N$ .

Factores de corrección de ganancia y desplazamiento para las imágenes de luminancia procesadas.

Información de registro espacial: desplazamiento horizontal y desplazamiento vertical. Para el vídeo entrelazado, el desplazamiento vertical de cada campo determina si el vídeo procesado requiere reencuadre.

Región válida de la secuencia de vídeo procesada (es decir PVR).

Incertidumbre ( $U$ ): Número que indica la precisión del registro temporal inicial. La hipótesis del registro temporal inicial es que el registro temporal auténtico para  $Y_P(t)$  esta dentro de más menos ( $U - HFw$ ) de  $Y_O(t)$ , para todo  $0 \leq t < N$ .

---

<sup>6</sup> Cuando el vídeo entrelazado requiere reencuadre, las longitudes de las secuencias de vídeo original y procesada deben reducirse en uno para acomodar el reencuadre. Esto reducirá la longitud del fichero en un cuadro de vídeo con respecto al  $N$  especificado en la Fig. 20.

### 6.4.1.3 Cuadros y campos

El algoritmo de registro temporal basado en cuadros funciona tanto para el vídeo progresivo como para el entrelazado. Si la secuencia de vídeo es progresiva, el algoritmo alinea cuadros. Si la secuencia de vídeo es entrelazada, el algoritmo alinea campos. Cuando se alinean secuencias de vídeo entrelazadas, se consideran cuadros o alineaciones reencuadradas pero no ambos. Cuando se consideran alineaciones de cuadros, el campo uno del vídeo procesado se compara con el campo uno del vídeo original, y el campo dos del vídeo procesado se compara con el campo dos del vídeo original. Cuando se consideran alineaciones reencuadradas, el campo uno del vídeo procesado se compara con el campo dos del vídeo original y el campo dos del vídeo procesado se compara con el campo uno del vídeo original. Los valores de registro espacial introducidos en el algoritmo determinan si se consideran alineaciones de cuadro o alineaciones reencuadradas. La presencia de reencuadre se detecta examinando el registro espacial vertical de cada campo. Si el desplazamiento vertical del campo uno es igual al desplazamiento vertical del campo dos, el vídeo procesado no está reencuadrado; sólo se consideran alineaciones de cuadro. Si el desplazamiento vertical del campo dos es una unidad mayor que el desplazamiento vertical del campo uno, sólo se consideran las alineaciones reencuadradas. Todas las demás combinaciones de desplazamientos verticales indican la existencia de problemas que deben resolverse antes del registro temporal.

### 6.4.1.4 Descripción del algoritmo

*Paso 1: Calibrar las secuencias de vídeo*

Corregir la secuencia de vídeo procesada,  $Y_P(t)$ , utilizando la información de registro espacial y ganancia-desplazamiento suministrada como entrada al algoritmo.

*Paso 2: Seleccionar la subregión de vídeo a utilizar*

La subregión de interés a utilizar por el algoritmo debe ser un múltiplo de BLOCK\_SIZE y debe caber en la PVR. Debe seleccionarse la mayor subregión que satisfaga estos dos requisitos y sea la más próxima al centro de la imagen. Todo el procesamiento restante se limitará al vídeo dentro de esta subregión de interés seleccionada.

*Paso 3: Submuestreo espacial de las imágenes original y procesada*

Submuestrear espacialmente la región de interés de  $Y_O(t)$  e  $Y_P(t)$  por un factor de BLOCK\_SIZE calculando la media de cada bloque. Para los cuadros de vídeo progresivo, el submuestreo será BLOCK\_SIZE horizontalmente y verticalmente, mientras que para los campos de vídeo entrelazado, el submuestreo será BLOCK\_SIZE horizontalmente y BLOCK\_SIZE/2 verticalmente. Por ejemplo, el submuestreo de una secuencia de vídeo progresivo por un BLOCK\_SIZE de 16 cogerá la media de cada bloque de 16 píxeles por 16 líneas de cuadro, mientras que el submuestreo de una secuencia de vídeo entrelazado por un BLOCK\_SIZE de 16 cogerá la media de cada bloque de 16 píxeles por 8 líneas de campo. Este submuestreo reduce la repercusión de las degradaciones en el proceso de registro temporal.

*Paso 4: Normalizar las imágenes submuestreadas*

Normalizar cada imagen submuestreada por la desviación típica de dicha imagen. No aplicar esta normalización a las imágenes cuya desviación típica sea inferior a uno (por ejemplo, aquellas que contienen un campo de color plano)<sup>7</sup>. Esta normalización minimizará la influencia de las

---

<sup>7</sup> La normalización se salta cuando la desviación típica es inferior a uno para evitar la amplificación del ruido y la posibilidad de dividir por cero en las imágenes que contengan un nivel de intensidad plano o uniforme.

fluctuaciones del contraste y la energía de las imágenes individuales en los resultados del registro temporal. Tras este Paso, las secuencias de vídeo original y procesado se representarán por  $\mathbf{S}_O(t)$  y  $\mathbf{S}_P(t)$ , respectivamente, para indicar que las imágenes se han submuestreado y normalizado.

*Paso 5: Comparar las imágenes procesadas con las originales*

Comparar cada imagen procesada,  $\mathbf{S}_P(t)$  con las imágenes originales  $\mathbf{S}_O(t+d)$ , siendo los valores válidos de  $d$ :  $(-U \leq d \leq +U)$  y los valores válidos de  $t$ :  $(U \leq t < N-U)$ . Para la imagen procesada  $t$  y la original  $t+d$ , estas comparaciones se indicarán por  $C_{td}$  y se calcularán como la desviación típica en el espacio de la imagen formada restando la imagen procesada  $t$  de la imagen original  $t+d$ :  $C_{td} = std_{space}(\mathbf{S}_O(t+d) - \mathbf{S}_P(t))$ . Estas comparaciones,  $C_{td}$ , correlacionan la  $t$ -ésima imagen procesada con cada imagen original que esté dentro de la incertidumbre de registro. Valores inferiores de  $C_{td}$  indican que la imagen procesada se parece más a la imagen original ya que es mayor la cancelación de la varianza de imagen. El rango de  $t$ ,  $U \leq t < N-U$ , cubre todas las imágenes procesadas para las que hay disponibles imágenes originales para todo el rango de incertidumbre del registro temporal.

*Paso 6: Efectuar una comprobación global de vídeo en reposo*

Para determinar si hay suficiente movimiento en la secuencia de vídeo, promediar  $C_{td}$  para el índice temporal  $t$  para cada  $d$ :

$$A_d = \frac{1}{N-2*U} \cdot \sum_{t=U}^{N-U-1} C_{td} \quad (100)$$

Esta suma (100) incluye el rango de imágenes de vídeo procesadas  $t$  para el que está disponible la incertidumbre completa de imágenes originales.  $A_d$  tiene un valor para cada retardo de registro temporal  $d$  considerado. Si  $(\text{máximo}(A_d) - \text{mínimo}(A_d) < \text{STILL\_THRESHOLD})$ , la escena no contiene suficiente movimiento para el registro temporal basado en cuadros. La escena completa está en reposo o prácticamente en reposo. Los resultados de la correlación de retardos de vídeo distintos son entonces tan parecidos que cualquier diferenciación se deberá a una casualidad aleatoria y no a mediciones fiables. Si se detecta una secuencia de vídeo en reposo, el usuario recibe un aviso a propósito y el algoritmo termina en este punto.

*Paso 7: Registrar temporalmente cada imagen procesada*

Para cada imagen procesada  $t$  ( $U \leq t < N-U$ ), encontrar la  $d$  dentro de la incertidumbre temporal  $(-U \leq d \leq +U)$  que minimiza  $C_{td}$ . Dicho de otro modo, para cada imagen procesada  $t$ , encontrar  $d_{min}(t)$  tal que  $C_{t \ d_{min}(t)} \leq C_{td}$ , para toda  $d$ . El mejor registro temporal de la imagen procesada  $t$  viene dado por  $d_{min}(t)$ . La mayor parte del tiempo, el registro temporal indicado para imágenes individuales será correcto o prácticamente correcto. El registro temporal será incorrecto para ciertas imágenes debido a diversos motivos (distorsión de la imagen, errores, ruido, movimiento insuficiente, etc.).

*Paso 8: Ejecutar una comprobación de reposo en cada imagen procesada*

Si para una determinada imagen  $t$  y todos los valores de  $d$  ( $-U \leq d \leq U$ ),  $\text{máximo}(C_{td}) - \text{mínimo}(C_{td}) < \text{STILL\_THRESHOLD}$ ,  $d_{min}(t)$  queda sin definir para esta imagen procesada  $t$ . En particular, no es suficiente el movimiento en torno a la imagen  $t$  para que el registro temporal basado en cuadros funcione adecuadamente.

*Paso 9: Construir un histograma de todos los registros temporales definidos*

Calcular un histograma con los valores definidos de  $d_{min}(t)$  con  $2*U+1$  acumuladores en el que cada acumulador representa un retardo de vídeo diferente (o sea, entre  $-U$  y  $+U$ ). Los valores de  $d_{min}(t)$  no definidos (por ejemplo las imágenes en reposo) quedan fuera del cálculo del histograma.

Este histograma, representado por  $H_d$ , es el histograma de retardos temporales de todas las imágenes procesadas que tienen suficiente movimiento para efectuar un registro temporal válido. Cada acumulador del histograma contiene el número de imágenes procesadas con dicho retardo de vídeo  $d$ , donde  $d$  puede tomar valores entre  $-U$  y  $+U$ .

*Paso 10: Formar un histograma suavizado*

El histograma  $H_d$  se suaviza por convolución del mismo con un filtro paso bajo de longitud  $2*HFW + 1$ , que se define para el índice  $k$  por:

$$F_k = \frac{0,5 + 0,5 * \cos[\pi * (k - HFW) / (1 + HFW)]}{\sum_{i=0}^{2*HFW} \{0,5 + 0,5 * \cos[\pi * (i - HFW) / (1 + HFW)]\}} \quad \text{para} \quad 0 \leq k \leq 2 * HFW \quad (101)$$

Cuando se considera el histograma suavizado (101),  $SH_d$ , que resulta de este Paso, los acumuladores HFW de cada extremo de  $SH_d$  se tratan como no definidos. Esto restringe los retardos de vídeo que pueden calcularse a más menos (UNCERTAINTY-HFW). El suavizado del histograma mejora la validez de las estimaciones del retardo de vídeo.

*Paso 11: Examinar la información del histograma*

A partir del histograma original,  $H_d$ , y del histograma suavizado,  $SH_d$ , se determinan los tres valores siguientes:

- max\_H\_value: Valor máximo de  $H_d$ .
- max\_SH\_offset: Desplazamiento  $d$  que maximiza  $SH_d$ .
- max\_SH\_value: Valor máximo de  $SH_d$  (por ejemplo a  $d = \text{max\_SH\_offset}$ ).

A continuación, se efectuarán las dos comprobaciones siguientes:

- ¿Ha sido  $U$  suficientemente grande? Recuérdese que el primer y el último acumulador HFW de  $H_d$  no están en  $SH_d$ . Examinar los valores de  $H_d$  en estos acumuladores. Si ( $H_d > \text{max\_H\_value} * \text{BELOW\_WARN}$ ), la incertidumbre del registro temporal es demasiado pequeña. El algoritmo debe volver a ejecutarse con una  $U$  superior. Los valores de  $d$  a comprobar son ( $-U \leq d < -U + HFW$ ) y ( $U - HFW < d \leq U$ ).
- ¿Tiene  $SH_d$  un retardo bien definido? Examinar  $SH_d$ , excepto dentro de DELTA de  $\text{max\_SH\_offset}$ . Si ( $SH_d > \text{max\_SH\_value} * \text{BELOW\_WARN}$ ) para cualquier retardo de vídeo  $d$  en el que ( $-U \leq d < \text{max\_SH\_offset} - \text{DELTA}$ ) o ( $\text{max\_SH\_offset} + \text{DELTA} < d \leq U$ ), el registro temporal es ambiguo.

Si se superan las dos comprobaciones anteriores, el retardo de vídeo dado por  $\text{max\_SH\_offset}$  se escoge como el mejor registro temporal medio para la escena.

#### 6.4.1.5 Observaciones y conclusiones

El algoritmo de medición del retardo de vídeo basado en cuadros utiliza secuencias de vídeo original y procesado submuestreadas. Este algoritmo es adecuado para la alineación de vídeo en un entorno fuera de servicio totalmente automatizado, antes de efectuar las mediciones de la calidad de vídeo. El algoritmo de medición de la calidad de vídeo basado en cuadros estima el registro temporal de cada imagen, construye histogramas de estas estimaciones individuales y, a continuación, utiliza el retardo indicado más frecuentemente como retardo global del vídeo, o registro temporal, para la secuencia seleccionada de cuadros de vídeo.

El retardo indicado en la etapa final del algoritmo (véase el Paso 11 del § 6.4.1.4) puede ser distinto del que podría escoger un espectador, si se alinean las escenas a ojo. Los espectadores tienden a centrarse en el movimiento, alineando las partes de mayor movimiento de la escena, en las que el

algoritmo basado en cuadros escoge el retardo observado más frecuentemente en todas las tramas examinadas. Pueden examinarse estos histogramas de retardo global para determinar la amplitud y estadística de cualquier retardo de vídeo variable presente en el HRC.

#### 6.4.2 Aplicación de la corrección del registro temporal

Todas las características de calidad requerirán la eliminación del retardo temporal aquí calculado. En los retardos positivos, eliminar los cuadros del principio del fichero procesado y del final del fichero original. En los retardos negativos, eliminar los cuadros del final del fichero procesado y del principio del fichero original. Cuando se reencuadran secuencias de vídeo entrelazadas, se reencuadra la secuencia procesada. Así pues, además de lo anterior, debe eliminarse un campo del principio de la secuencia de vídeo procesada y otro del final. Simultáneamente, debe eliminarse un cuadro ya sea del principio del fichero de vídeo original (es decir  $-1$  campo de retardo global) o del final de fichero de vídeo original (es decir  $+1$  campo de retardo global).

La corrección del registro temporal acortará de hecho la longitud de las imágenes disponibles de la secuencia de vídeo. Para mayor simplicidad, todos los cálculos adicionales se basarán en el número de cuadros de vídeo disponibles una vez aplicadas las correcciones de calibración.

### 7 Características de calidad

#### 7.1 Introducción

Una característica de calidad se define como la cantidad de información asociada a una subregión espacio-temporal de una secuencia de vídeo (original o procesada) o extraída de ésta. Las secuencias de características producidas son función del espacio y del tiempo. Comparando las características extraídas del vídeo procesado calibrado con las extraídas del vídeo original, puede calcularse un conjunto de parámetros de calidad (véase el § 8) indicativos de las variaciones perceptuales de la calidad de vídeo. En este apartado se describe un conjunto de características de calidad que caracterizan las variaciones perceptuales de las propiedades espaciales, temporales y de crominancia de las secuencias de vídeo. Normalmente, se aplica un filtro perceptual a la secuencia de vídeo para mejorar alguna propiedad de la calidad del vídeo percibido, tal como la información de los bordes. Tras este filtrado perceptual, se extraen las características de las subregiones espacio-temporales (S-T) utilizando una función matemática (por ejemplo la desviación típica). Por último, se aplica un umbral de perceptibilidad a las características extraídas.

En el texto que sigue, una secuencia de características originales se representa por  $f_o(s, t)$  y la correspondiente secuencia de características procesadas se representa por  $f_p(s, t)$ , siendo  $s$  y  $t$  los índices que representan las posiciones espacial y temporal, respectivamente de la región S-T dentro de las secuencias de vídeo original y procesada calibrada. Se asignarán a las características subíndices de letras conforme aparecen en los apartados siguientes, escogiéndose las letras de los subíndices de modo que representen lo que las características miden. Todas las características operan sobre cuadros de una secuencia de vídeo calibrada (véase el § 6); las posibles cuestiones de entrelazado se abordan durante la calibración. Todas las características operan con independencia del tamaño de la imagen (es decir, el tamaño de la región S-T no varía al hacerlo el tamaño de la imagen)<sup>8</sup>.

---

<sup>8</sup> Existe la hipótesis implícita de que la distancia de observación como función de la altura de la imagen permanece fija (o sea, se utilizan distancias de observación menores para las imágenes más pequeñas). Pueden verse en el § 9 otras observaciones relativas a la distancia de observación supuesta.



En resumen, los cálculos de las características se realizan de acuerdo con los siguientes Pasos, aunque algunas características tal vez no requieran los Pasos marcados como [Opcional].

*Paso 1:* [Opcional] Aplicar un filtro perceptual.

*Paso 2:* Dividir la secuencia de vídeo en regiones S-T.

*Paso 3:* Extraer las características, o estadísticas resumen, de cada región S-T (por ejemplo, media y desviación típica).

*Paso 4:* [Opcional] aplicar un umbral de perceptibilidad.

Determinadas características pueden utilizar dos o más filtros perceptuales distintos.

### 7.1.1 Regiones S-T

En general, las características se extraen de regiones S-T localizadas, una vez filtradas perceptualmente las secuencias de vídeo original y procesada. La situación de las regiones S-T corresponde a la división de las secuencias de vídeo en regiones S-T adyacentes. Como el vídeo procesado se ha calibrado, para cada región S-T de vídeo procesado existe una región S-T original que ocupa la misma posición espacio-temporal dentro de la secuencia de vídeo. Las características se extraen de cada región S-T calculando las estadísticas resumen o alguna otra función matemática sobre la región S-T de interés.

Cada región S-T describe un bloque de píxeles. Los tamaños de las regiones S-T se describen por:

- el número de píxeles horizontalmente,
- el número de líneas de cuadro verticalmente, y
- la duración temporal de la región, expresada en unidades de cuadros de vídeo equivalentes con referencia a un sistema de vídeo de 30 fps<sup>9</sup>.

En la Fig. 27 se representa una región S-T de 8 píxeles horizontales  $\times$  8 líneas verticales  $\times$  6 cuadros de vídeo NTSC, correspondientes a un total de 384 píxeles. Cuando se aplica a vídeo de 25 fps (PAL), esta misma región S-T abarca 8 píxeles horizontales  $\times$  8 líneas verticales  $\times$  5 cuadros de vídeo, lo que corresponde a un total de 320 píxeles.

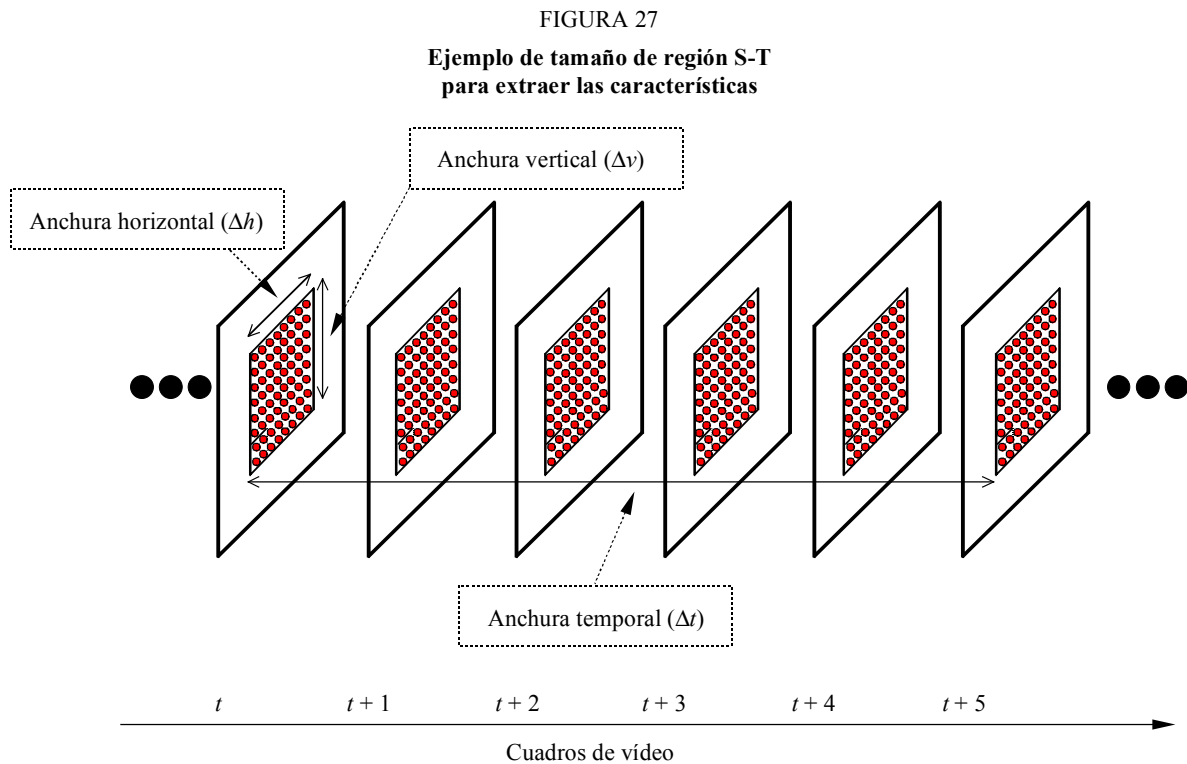
Un quinto de segundo es una amplitud temporal conveniente, debido a la facilidad de conversión de la velocidad de cuadros (es decir un quinto de segundo se traduce en un número entero de cuadros de vídeo en los sistemas de vídeo que funcionan a 10 fps, 15 fps, 25 fps y 30 fps). La regla general para la conversión de la velocidad de cuadros consiste en tomar la longitud de la región S-T en las secuencias de vídeo de 30 fps, dividirla por 30 y multiplicarla por la velocidad de cuadros del sistema de vídeo objeto de la prueba. Se supone que las regiones S-T que contengan un cuadro de vídeo contendrán siempre un cuadro de vídeo, con independencia de la velocidad de cuadros.

La región espacial de interés (SROI, véase el § 3) ocupada por todas las regiones S-T es idéntica para las secuencias de vídeo original y procesada calibrada. La SROI debe quedar totalmente dentro de la PVR, posiblemente con una zona intermedia de píxeles necesaria para cualquier filtro

---

<sup>9</sup> Todas las duraciones temporales del presente Anexo estarán referidas al número equivalente de cuadros de vídeo de un sistema de vídeo de 30 fps. Por consiguiente, la duración temporal de 6 cuadros ( $F$ ) se utiliza para representar tanto los 6 cuadros de un sistema NTSC ( $6/30$ ) como los 5 cuadros de un sistema PAL ( $5/25$ ). Además, 30 fps y 29,97 fps se utilizan indistintamente en este Anexo, ya que la pequeña diferencia en velocidad de cuadros no tiene repercusión alguna en el cálculo de la VQM.

perceptual convolucional. La anchura horizontal de la SROI debe ser divisible exactamente por la amplitud horizontal de la región S-T. Análogamente, la altura vertical de la SROI debe ser divisible exactamente por la amplitud vertical de la región S-T. El usuario podría imponer una restricción adicional sobre la SROI para ocupar una región de interés particular tal como el centro del cuadro de vídeo.



1683-27

Temporalmente, las secuencias de vídeo original y procesada calibrada se dividen en un número idéntico de regiones S-T, comenzando por el primer cuadro de vídeo alineado temporalmente. Si el número de cuadros válidos disponibles no pudiera dividirse exactamente por la amplitud temporal de la región S-T, se suprimirían del estudio los cuadros del final del corto.

Para algunas características tales como las presentadas en el § 7.2, el bloque de  $8 \times 8$ \_6F consigue una correlación próxima a la máxima con las valoraciones subjetivas. Debe observarse, no obstante, que la correlación disminuye lentamente al alejarse del tamaño óptimo de la región S-T. Pueden utilizarse anchuras horizontales y verticales de hasta 32 o incluso superiores, y anchuras temporales de hasta 30 cuadros con resultados satisfactorios lo que ofrece al diseñador del sistema de medición objetiva bastante flexibilidad para adaptar las características al almacenamiento disponible o a la anchura de banda de transmisión [Wolf y Pinson, 2001].

Una vez dividida la secuencia de vídeo en regiones S-T, el eje temporal de la característica ( $t$ ) ya no corresponde a cuadros individuales. Por el contrario, el eje temporal contiene un número de muestras igual al número de cuadros válidos de la secuencia de vídeo calibrada, dividido por la amplitud temporal de la región S-T.

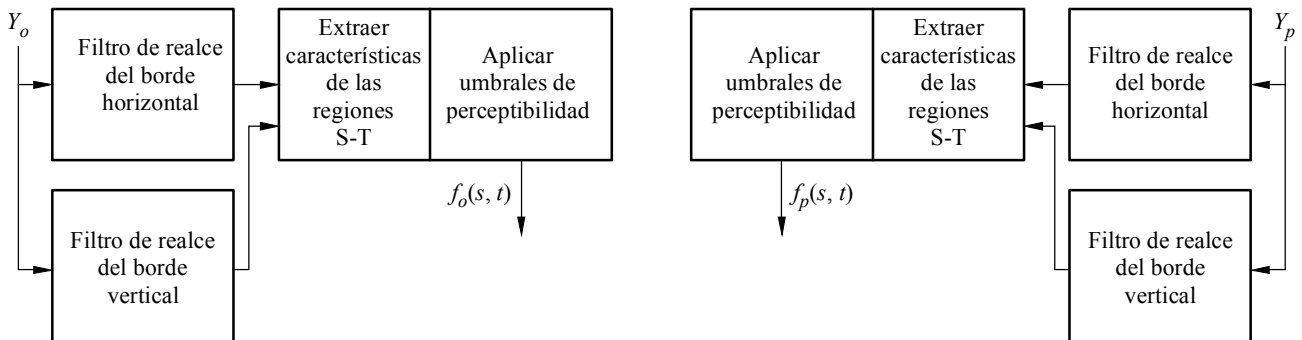
Cuando se calculan dos o más características simultáneamente, las consideraciones adicionales cobran importancia. Teóricamente, todas las características deben calcularse para la misma SROI.

## 7.2 Características basadas en gradientes espaciales

Se pueden utilizar características derivadas de gradientes espaciales para caracterizar las distorsiones perceptuales de los bordes. Por ejemplo, el difuminado provoca una pérdida general de información de los bordes mientras que un exceso de información de los bordes horizontales y verticales puede provocar la distorsión de bloque o entejamiento. Las componentes  $Y$  de las secuencias de vídeo original y procesada se filtran por medio de filtros de realce de bordes horizontales y verticales. A continuación, estas secuencias de vídeo filtradas se dividen en regiones S-T de las que se extraen las características o estadísticas resumen, que cuantifican la actividad espacial en función de la orientación angular. A continuación, las características se recortan en el extremo inferior para emular los umbrales de perceptibilidad. Los filtros de realce de bordes, el tamaño de la región S-T, y los umbrales de perceptibilidad se seleccionaron con arreglo al vídeo de la Recomendación UIT-R BT.601, evaluados subjetivamente a una distancia de observación de seis alturas de imagen. En la Fig. 28 se presenta un esquema del algoritmo utilizado para extraer las características basadas en gradientes espaciales.

FIGURA 28

Esquema del algoritmo utilizado para extraer las características del gradiente espacial



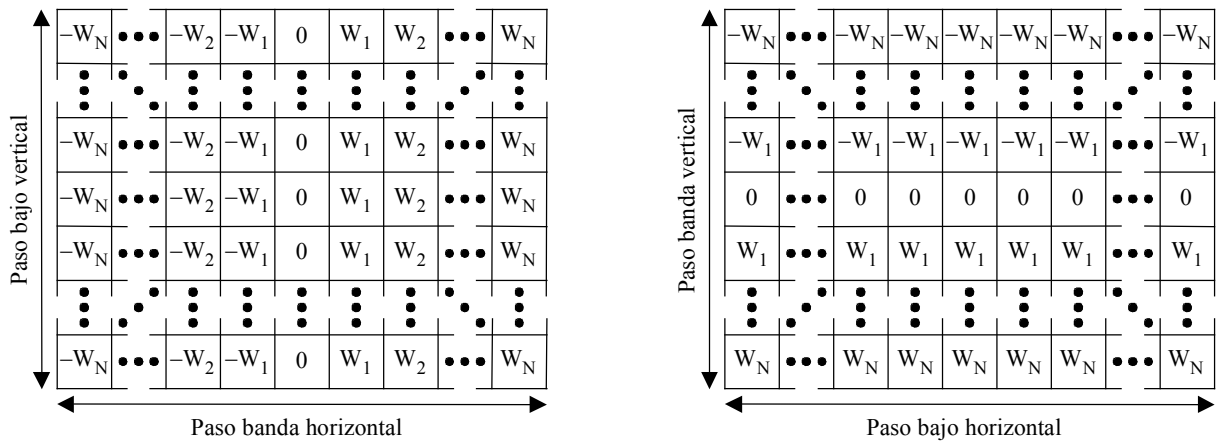
1683-28

### 7.2.1 Filtros de realce de bordes

Los cuadros de vídeo  $Y$  (de luminancia) original y procesado se procesan primero con filtros de realce de bordes horizontal y vertical, que mejoran los bordes y reducen el ruido. Los dos filtros representados en la Fig. 29 se aplican por separado, uno para mejorar las diferencias de los píxeles horizontales suavizando al mismo tiempo en sentido vertical (filtro izquierdo), y el otro mejorando las diferencias de los píxeles verticales suavizando al mismo tiempo en sentido horizontal (filtro derecho).

FIGURA 29

## Filtros de realce de bordes



1683-29

Los dos filtros son transposiciones recíprocas, de tamaño  $13 \times 13$ , y con factores de ponderación dados por:

$$w_x = k \cdot \left(\frac{x}{c}\right) \cdot \exp\left\{-\left(\frac{1}{2}\right)\left(\frac{x}{c}\right)^2\right\}$$

siendo:

- $x$ : desplazamiento en píxeles del centro del filtro (0, 1, 2, ...,  $N$ )
- $c$ : constante que determina la anchura del filtro paso banda
- $k$ : constante de normalización seleccionada de modo que cada filtro produzca la misma ganancia como un auténtico filtro Sobel [Jain, 1989].

La magnitud óptima de filtrado de paso banda horizontal correspondiente a una distancia de observación de seis veces la altura de la imagen se obtuvo con el filtro  $c=2$ , que tiene una respuesta de cresta a 4,5 ciclos/grado aproximadamente. Los factores de ponderación del filtro paso banda utilizado son los siguientes:

$$[-0,0052625; -0,0173446; -0,0427401; -0,0768961; -0,0957739; -0,0696751; 0; 0,0696751; 0,0957739; 0,0768961; 0,0427401; 0,0173446; 0,0052625]$$

Obsérvese que los filtros de la Fig. 29 tienen una respuesta de paso bajo plana. La respuesta de paso bajo plana produjo la mejor estimación de calidad y tiene la ventaja adicional de ser computacionalmente eficiente (por ejemplo, para el filtro de la izquierda en la Fig. 29, sólo es necesario sumar los píxeles de una columna y multiplicar una vez por el factor de ponderación).

### 7.2.2 Descripción de las características $f_{S113}$ y $f_{HV13}$

En este apartado se describe la extracción de dos características de actividad espacial de las regiones S-T de las secuencias de vídeo original y procesada de bordes mejorados del § 7.2.1. Se utilizarán estas características para detectar degradaciones espaciales tales como la borrosidad y el bloqueo. El filtro representado en la Fig. 29 (a la izquierda) realza los gradientes espaciales en sentido horizontal,  $H$ , mientras que el transpuesto de este filtro (a la derecha) realza los gradientes

espaciales en sentido vertical, V. La respuesta en cada píxel de los filtros  $H$  y  $V$  puede dibujarse en un diagrama bidimensional tal como el representado en la Fig. 30 con la respuesta del filtro  $H$  en abscisas y la del filtro  $V$  en ordenadas. Para un determinado píxel de imagen situado en la fila  $i$ , columna  $j$ , e instante  $t$ , las respuestas del filtro  $H$  y  $V$  se representarán por  $H(i, j, t)$  y  $V(i, j, t)$ , respectivamente. Estas respuestas pueden pasarse a coordenadas polares  $(R, \theta)$  por medio de las relaciones siguientes:

$$R(i, j, t) = \sqrt{H(i, j, t)^2 + V(i, j, t)^2}$$

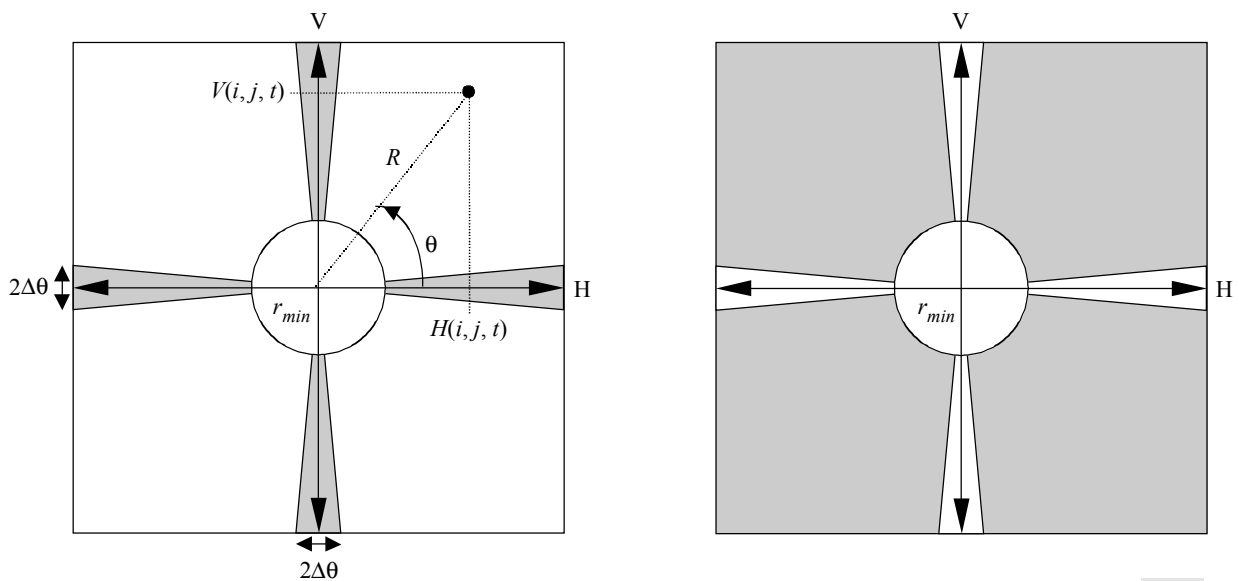
y

(102)

$$\theta(i, j, t) = \text{tg}^{-1} \left[ \frac{V(i, j, t)}{H(i, j, t)} \right]$$

FIGURA 30

División de la actividad espacial horizontal (H) y vertical (V) en distribuciones HV (izquierda) y HV (derecha)



1683-30

La primera característica es una medida de la SI global y por consiguiente se representa por  $f_{SI13}$ , ya que las imágenes se preprocesaron con las máscaras de  $13 \times 13$  filtros representadas en la Fig. 29. Esta característica se calcula sencillamente como desviación típica (std, *standard deviation*) sobre la región S-T de las muestras  $R(i, j, t)$ , y a continuación se recorta hasta el umbral de perceptibilidad de  $P$  (es decir si el resultado del cálculo de std es inferior a  $P$ ,  $f_{SI13}$  se hace igual a  $P$ ), o sea:

$$f_{SI13} = \{ \text{std}[R(i, j, t)] \}_P / i, j, t \in \{ \text{S-T region} \} \quad (103)$$

Esta característica es sensible a las variaciones de la magnitud global de actividad espacial de una determinada región S-T. Por ejemplo, la borrosidad localizada produce una reducción de la magnitud de actividad espacial, mientras que el ruido produce un aumento. El umbral  $P$  recomendado para esta característica es 12.

La segunda característica,  $f_{HV13}$ , es insensible a las variaciones de la distribución angular, u orientación, de la actividad espacial. Las imágenes complementarias se calculan con las distribuciones de gradiente espacial sombreadas representadas en la Fig. 30. La imagen con gradientes horizontales y verticales, representada por  $HV$ , contiene los píxeles  $R(i, j, t)$  que son bordes horizontales o verticales (los píxeles que son en bordes diagonales se anulan). La imagen con los gradientes diagonales, representada por  $\overline{HV}$ , contiene los píxeles  $R(i, j, t)$  que son bordes diagonales (los píxeles que son bordes horizontales o verticales se anulan). Las magnitudes de los gradientes  $R(i, j, t)$  menores que  $r_{min}$  se anulan en ambas imágenes para garantizar la exactitud de los cálculos de  $\theta$ . Los píxeles de  $HV$  y  $\overline{HV}$  pueden representarse matemáticamente del siguiente modo:

$$HV(i, j, t) = \begin{cases} R(i, j, t) & \text{si } R(i, j, t) \geq r_{min} \text{ y } m\frac{\pi}{2} - \Delta\theta < \theta(i, j, t) < m\frac{\pi}{2} + \Delta\theta \quad (m = 0, 1, 2, 3) \\ 0 & \text{en los demás casos} \end{cases} \quad (104)$$

y

$$\overline{HV}(i, j, t) = \begin{cases} R(i, j, t) & \text{si } R(i, j, t) \geq r_{min} \text{ y } m\frac{\pi}{2} + \Delta\theta \leq \theta(i, j, t) \leq (m+1)\frac{\pi}{2} - \Delta\theta \quad (m = 0, 1, 2, 3) \\ 0 & \text{en los demás casos} \end{cases} \quad (105)$$

donde:

$$i, j, t \in \{\text{S-T region}\}$$

Para el cálculo de las mencionadas  $HV$  y  $\overline{HV}$  el valor recomendado para  $r_{min}$  es 20 y para  $\Delta\theta$  0,225 radianes. Así pues, la característica  $f_{HV13}$  para una región S-T viene dada por la relación de la media de  $HV$  a la media de  $\overline{HV}$ , donde las medias resultantes se recortan en sus umbrales de perceptibilidad  $P$ , es decir:

$$f_{HV13} = \frac{\{\text{mean}[HV(i, j, t)]\}_P}{\{\text{mean}[\overline{HV}(i, j, t)]\}_P} \quad (106)$$

El umbral de perceptibilidad  $P$  recomendado para la media de  $HV$  y  $\overline{HV}$  es 3. La característica  $f_{HV13}$  es sensible a las variaciones de distribución angular de la actividad espacial de una determinada región S-T. Por ejemplo, si los bordes horizontales y verticales son más borrosos que los bordes diagonales,  $f_{HV13}$  del vídeo procesado será inferior a  $f_{HV13}$  del vídeo original. Por otra parte, si se introducen bordes horizontales o verticales erróneos, por ejemplo en forma de bloqueo o de distorsiones de entejado,  $f_{HV13}$  del vídeo procesado será mayor que  $f_{HV13}$  del vídeo original. Así pues la característica  $f_{HV13}$  proporciona un medio sencillo de tener en cuenta las variaciones de sensibilidad del sistema visual humano con respecto a la orientación angular<sup>10</sup>.

<sup>10</sup> Esta exposición de  $f_{HV13}$ , aunque es cierta en general, resulta un tanto simplificada. Por ejemplo, cuando se encuentran determinadas formas, el comportamiento del filtro de  $f_{HV13}$  puede parecer un contrasentido (por ejemplo, una esquina formada por la unión de una línea vertical y otra horizontal se traducirá en energía diagonal).

### 7.3 Características basadas en la información de crominancia

En este apartado se presenta una característica única que puede utilizarse para medir las distorsiones de las señales de crominancia ( $C_B$ ,  $C_R$ ). Para un determinado píxel de imagen situado en la fila  $i$ , columna  $j$  e instante  $t$ , sean  $C_B(i, j, t)$  y  $C_R(i, j, t)$  las representaciones de los valores  $C_B$  y  $C_R$  de la Recomendación UIT-R BT.601<sup>11</sup>. Las componentes de un vector de característica de crominancia bidimensional,  $f_{COHER\_COLOR}$ , se calculan como media, mean, en la región S-T de las muestras  $C_B(i, j, t)$  y  $C_R(i, j, t)$ , respectivamente, dando más peso perceptual a la componente  $C_R$ :

$$f_{COHER\_COLOR} = (\text{mean}[C_B(i, j, t)], W_R * \text{mean}[C_R(i, j, t)]) / i, j, t \in \{\text{S-T region}\}, \text{ y } W_R = 1,5 \quad (107)$$

La ecuación (107) realiza una integración coherente (de aquí el nombre  $f_{COHER\_COLOR}$ ) ya que la relación de fases entre  $C_B$  y  $C_R$  se conserva. Si se está familiarizado con el vectorscopio, el valor de la característica de crominancia cuando se examinan las señales de barra de color es fácilmente observable. Para escenas polivalentes, es posible imaginar la utilidad del vector de la característica de crominancia para medir distorsiones de crominancia en bloques de vídeo que cubran un rango de amplitudes espaciales y temporales. No obstante, si el tamaño de la región S-T es demasiado grande, pueden incluirse muchos colores en el cálculo, y la utilidad de  $f_{COHER\_COLOR}$  es menor. Un tamaño de región S-T de 8 píxeles horizontales  $\times$  8 líneas verticales  $\times$  (1 a 3) cuadros de vídeo produce un vector de característica de crominancia válido (realmente 4 píxeles horizontales  $C_B$  y  $C_R$ , ya que estas señales son submuestreadas por dos en sentido horizontal con arreglo al muestreo de la Recomendación UIT-R BT.601).

### 7.4 Características basadas en la información de contraste

Las características que miden información de contraste localizada son sensibles a degradaciones de calidad tales como la borrosidad (por ejemplo, pérdida de contraste) y el ruido añadido (por ejemplo, ganancia de contraste). Una característica de contraste localizada,  $f_{CONT}$ , se calcula fácilmente para cada región S-T de la imagen de luminancia  $Y$  del siguiente modo:

$$f_{CONT} = \{\text{std}[Y(i, j, t)]\}_P / i, j, t \in \{\text{S-T region}\} \quad (108)$$

El umbral de perceptibilidad  $P$  recomendado para la característica  $f_{CONT}$  se encuentra entre cuatro y seis.

### 7.5 Características basadas en ATI

Las características que miden distorsiones en el flujo de movimiento son sensibles a degradaciones de calidad tales como la omisión o repetición de cuadros (pérdida de movimiento) y la adición de ruido (ganancia de movimiento). Se calcula una característica de ATI,  $f_{ATI}$ , para cada región S-T generando primero una secuencia de vídeo en movimiento que sea el valor absoluto de la diferencia entre los cuadros de vídeo consecutivos en los instantes  $t$  y  $t - 1$ , y calculando a continuación la desviación típica en la región S-T. La representación matemática de este proceso es la siguiente:

$$f_{ATI} = \{\text{std}|Y(i, j, t) - Y(i, j, t - 1)|\}_P / i, j, t \in \{\text{S-T region}\} \quad (109)$$

El umbral de perceptibilidad recomendado  $P$  para la característica  $f_{ATI}$  está entre uno y tres.

---

<sup>11</sup> Las correcciones de la ganancia y del desplazamiento no se aplican a los planos de imagen  $C_B$  y  $C_R$ . Véase el § 6.3.3.

La utilización de un cuadro previo introduce consideraciones más allá de las exigidas por las otras características. Cuando se calcula  $f_{ATI}$  junto con otra característica (por ejemplo  $f_{CONTRAST\_ATI}$  del § 7.6) o para utilizarla en un modelo (véase el § 9), el requisito de un cuadro suplementario complica la tarea de emplazamiento de las regiones S-T (véase el § 7.1.1).

## 7.6 Características basadas en el producto vectorial del contraste y la ATI

La perceptibilidad de las degradaciones espaciales puede verse influida por la cantidad de movimiento presente. Análogamente, la perceptibilidad de las degradaciones temporales puede verse influida por la cantidad de detalle espacial presente. Puede utilizarse una característica derivada del producto vectorial de la información de contraste y de la información temporal absoluta para explicar, al menos en parte, estas interacciones. Esta característica, representada por  $f_{CONTRAST\_ATI}$ , se calcula como producto de las características de los § 7.4 y 7.5<sup>12</sup>. El umbral de perceptibilidad recomendado  $P = 3$  se aplica a cada característica ( $f_{CONT}$  y  $f_{ATI}$ ) por separado, antes de calcular su producto vectorial. Las degradaciones serán más visibles en las regiones S-T cuyo producto vectorial sea bajo que en aquéllas cuyo producto vectorial sea alto. Esto es especialmente cierto para degradaciones como el ruido y los bloques con errores.

El requisito de un cuadro suplementario para  $f_{ATI}$  complica  $f_{CONTRAST\_ATI}$  ligeramente ya que las regiones S-T utilizadas por  $f_{CONT}$  y  $f_{ATI}$  deben estar situadas idénticamente. O bien debe quedar sin utilizar un cuadro al principio de la secuencia de vídeo para  $f_{ATI}$ , o las regiones S-T situadas al principio de la secuencia de vídeo deben contener un cuadro menos (por ejemplo, dada una amplitud temporal de  $6F$ , la primera región S-T de  $f_{ATI}$  utilizaría  $5F$  en vez de  $6F$ ). Los parámetros y modelos aquí especificados suponen la utilización de la segunda solución.

## 8 Parámetros de calidad

### 8.1 Introducción

Los parámetros de calidad que miden la distorsión de la calidad de vídeo debida a las pérdidas y ganancias de los valores de las características se calculan en primer lugar para cada región S-T comparando los valores originales de las características,  $f_o(s, t)$ , con los correspondientes valores de las características procesados,  $f_p(s, t)$  (véase el § 8.2). Para emular el enmascaramiento visual de las degradaciones en cada región S-T se utilizan varias relaciones funcionales. A continuación, se emula mediante funciones de combinación de errores en el espacio y el tiempo, la manera en que las personas deducen los índices de calidad subjetivos. La combinación de errores en el espacio se denominará agrupamiento espacial (véase el § 8.3), y la combinación de errores en el tiempo se denominará agrupamiento temporal (véase el § 8.4). La aplicación secuencial de las funciones de agrupamiento espacial y temporal a la secuencia de parámetros de calidad S-T produce parámetros de calidad para todo el corto de vídeo, cuya duración nominal es de 5 a 10 s. Los valores finales del parámetro agrupado en el tiempo pueden ajustarse y recortarse (véase el § 8.5) en consideración de la relación no lineal entre el valor del parámetro y la calidad percibida, y además para reducir la sensibilidad del parámetro.

---

<sup>12</sup> Se utiliza un producto vectorial estándar de las características  $f_{CONT}$  y  $f_{ATI}$  (es decir  $f_{CONT} * f_{ATI}$ ) para las características  $f_p(s, t)$  procesada y  $f_o(s, t)$  original en las funciones de comparación `ratio_loss` y `ratio_gain` descritas en el § 8.2.1. No obstante, para las funciones de comparación `log_loss` y `log_gain` las características procesada y original se calculan como  $\log_{10}[f_{CONT}] * \log_{10}[f_{ATI}]$ , y las funciones de comparación utilizan la resta (es decir  $f_p(s, t) - f_o(s, t)$  en vez de  $\log_{10}[f_p(s, t)] / \log_{10}[f_o(s, t)]$ ).



En resumen, el cálculo de los parámetros se ajusta a los siguientes Pasos. Ciertas características tal vez no requieran los Pasos [Opcional].

*Paso 1:* Comparar los valores originales de la característica con los valores de la característica procesados.

*Paso 2:* Ejecutar el agrupamiento espacial.

*Paso 3:* Ejecutar el agrupamiento temporal.

*Paso 4:* [Opcional] ejecutar el recorte y/o ajuste no lineal.

Los parámetros se diseñan de modo que sean todos positivos o todos negativos. Un parámetro cuyo valor sea cero indica que no hay degradación.

## 8.2 Funciones de comparación

La degradación perceptual en cada región S-T se calcula por medio de funciones que modelan el enmascaramiento visual de las degradaciones espaciales y temporales. En este apartado se presentan las funciones de enmascaramiento utilizadas por los diversos parámetros para producir parámetros de calidad en función del espacio y del tiempo.

### 8.2.1 Tasa de errores y tasa logarítmica

La atenuación y la ganancia se suelen examinar por separado, ya que producen esencialmente efectos distintos sobre la percepción de la calidad (por ejemplo, pérdida de la actividad espacial debida a la borrosidad y ganancia de la actividad espacial debida al ruido o al bloqueo). Hay dos formas, de entre las muchas funciones de comparación evaluadas, que han producido constantemente la mejor correlación con los índices subjetivos. Pueden utilizarse estas formas ya sea en los cálculos de la ganancia o en los de la atenuación para un total de cuatro funciones básicas de comparación S-T. Las cuatro formas primarias son las siguientes:

$$\text{ratio\_loss}(s,t) = np \left\{ \frac{f_p(s,t) - f_o(s,t)}{f_o(s,t)} \right\}$$

$$\text{ratio\_gain}(s,t) = pp \left\{ \frac{f_p(s,t) - f_o(s,t)}{f_o(s,t)} \right\}$$

$$\text{log\_loss}(s,t) = np \left\{ \log_{10} \left[ \frac{f_p(s,t)}{f_o(s,t)} \right] \right\}$$

$$\text{log\_gain}(s,t) = pp \left\{ \log_{10} \left[ \frac{f_p(s,t)}{f_o(s,t)} \right] \right\}$$

siendo:

*pp*: operador parte positiva (es decir los valores negativos se sustituyen por cero)

*np*: operador parte negativa (es decir los valores positivos se sustituyen por cero).

Estas funciones de enmascaramiento visual suponen que la percepción de la degradación es inversamente proporcional a la cantidad de actividad espacial o temporal localizada presente. Dicho de otro modo, las degradaciones espaciales resultan menos visibles conforme aumenta la actividad espacial (es decir el enmascaramiento espacial), mientras que las degradaciones temporales se vuelven menos visibles conforme aumenta la actividad temporal (es decir el enmascaramiento temporal). Mientras que la funciones de comparación logarítmica y la tasa se comportan de un modo muy parecido, la función logarítmica tiende a favorecer ligeramente a las ganancias mientras que la función tasa tiende a favorecer ligeramente a la atenuación. La función logarítmica tiene un margen dinámico superior y es útil cuando los valores procesados de la característica superan con mucho a los valores originales de la característica.

### 8.2.2 Distancia euclidiana

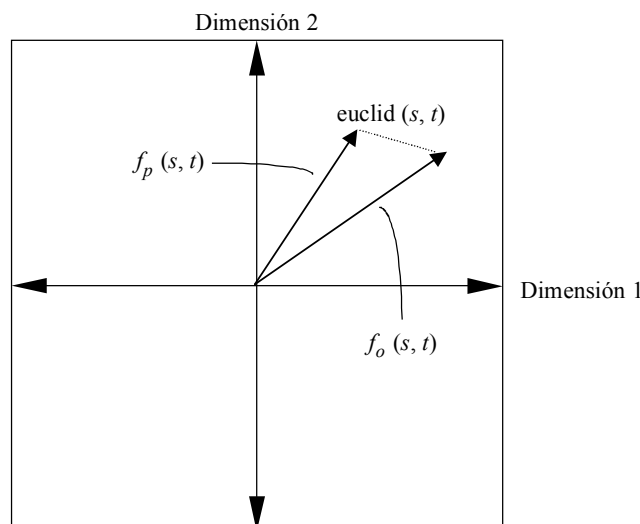
Otra función de comparación S-T de gran utilidad es la simple distancia euclidiana, representada por la longitud del vector diferencia entre el vector original de la característica  $f_o(s,t)$  y el correspondiente vector procesado de la característica,  $f_p(s,t)$ :

$$\text{euclid}(s,t) = \left\| \underline{f}_p(s,t) - \underline{f}_o(s,t) \right\| \quad (110)$$

La Fig. 31 representa la distancia euclidiana correspondiente a un vector de característica bidimensional extraído de una región S-T (por ejemplo, el vector de la característica  $f_{\text{COHER\_COLOR}}$  del § 7.3), siendo  $s$  y  $t$  índices que representan las posiciones espacial y temporal, respectivamente, de la región S-T en las secuencias de vídeo original y procesada calibrada. La línea de puntos de la Fig. 31 es la distancia euclidiana. La medida de la distancia euclidiana puede generalizarse a vectores de características que tengan un número arbitrario de dimensiones.

FIGURA 31

Representación de la distancia euclidiana  $\text{euclid}(s,t)$  correspondiente a un vector de característica bidimensional



### 8.3 Funciones de agrupamiento espacial

Los parámetros de las regiones S-T (véase el § 8.2) forman matrices tridimensionales que comprenden un eje temporal y dos dimensiones espaciales (es decir, el emplazamiento horizontal y vertical de la región S-T). A continuación, se agrupan las degradaciones de las regiones S-T que tengan el mismo índice  $t$  utilizando una función de agrupamiento espacial. El agrupamiento espacial produce una historia temporal de los valores de los parámetros. Esta historia temporal de los valores de los parámetros, representada genéricamente por  $p(t)$ , debe, a continuación, agruparse temporalmente por medio de una función de agrupamiento temporal definida en el § 8.4. El Cuadro 12 presenta un resumen de las funciones de agrupamiento espacial más utilizadas.

Se han realizado investigaciones exhaustivas que han puesto de manifiesto que las funciones de agrupamiento espacial óptimas suelen conllevar algún tipo de procesamiento correspondiente al caso más desfavorable, tal como el promedio del 5% más desfavorable de las distorsiones observadas en el índice espacial  $s$  [Wolf y Pinson, 1998, 1999, 2001 y 2002]. Esto se debe a que las degradaciones localizadas tienden a atraer la atención del observador, haciendo que la parte más desfavorable de la imagen sea el factor predominante en la decisión de calidad subjetiva. Por ejemplo, la función de agrupamiento espacial «above95%» (por encima del 95%) se calcula en cada índice temporal  $t$  para la función  $\log\_gain(s,t)$  de § 8.2.1 como el promedio del 5% más positivo de los valores sobre el índice espacial  $s$ <sup>13</sup>. Esto equivale a ordenar las distorsiones de la ganancia de menor a mayor en cada índice temporal  $t$  y promediar aquellas que se encuentran por encima del umbral del 95% (ya que valores más positivos supondrían mayor distorsión). Análogamente, las distorsiones de atenuación tales como las producidas por la función  $ratio\_loss(s,t)$  de § 8.2.1 se ordenarían en cada índice temporal  $t$ , aunque se utilizaría el promedio de las distorsiones «below5%» (por debajo del 5%) (ya que las atenuaciones son negativas).

### 8.4 Funciones de agrupamiento temporal

A continuación los resultados de la historia temporal del parámetro  $p(t)$ , obtenidos de la función de agrupación espacial (véase el § 8.3), se agrupan por medio de una función de agrupamiento temporal para producir un parámetro objetivo  $p$  para el corto de vídeo, que tiene un valor nominal de 4 a 10 s de longitud. Los observadores pueden utilizar varias funciones de agrupamiento temporal cuando estén calificando subjetivamente cortos de vídeo que tengan una duración aproximada de 10 s. La media en el tiempo es indicativa de la calidad promedio observada durante el periodo de tiempo. Los niveles 90% y 10% en el tiempo son indicativos de la peor calidad transitoria observada para las ganancias y atenuaciones, respectivamente (por ejemplo, los errores de transmisión digital pueden provocar una perturbación de 1 a 2 s en el vídeo procesado). Tras el agrupamiento temporal, un determinado parámetro  $p$  es o bien totalmente negativo o bien totalmente positivo, aunque no ambos. El Cuadro 13 presenta un resumen de las funciones de agrupamiento temporal más utilizadas.

---

<sup>13</sup> Obsérvese que el índice temporal,  $t$ , no indica aquí cuadros individuales (véase el § 7.1.1). En cambio, cada valor de  $t$  corresponde a regiones S-T con la misma amplitud temporal.

CUADRO 12

**Funciones de agrupamiento espacial y definiciones**

<b>Función de agrupamiento espacial</b>	<b>Definición</b>
below5% (por debajo del 5%)	Para cada índice temporal $t$ , ordenar los valores de los parámetros de menor a mayor. Calcular el promedio de todos los valores del parámetro que sean menores o iguales que el nivel umbral 5%. Para los parámetros de atenuación, esta función de agrupamiento espacial produce un parámetro indicativo de la calidad más desfavorable en el espacio
above95% (por encima del 95%)	Para cada índice temporal $t$ , ordenar los valores del parámetro de menor a mayor. Calcular el promedio de todos los valores de parámetro que sean mayores o iguales que el 95% del nivel umbral. Para los parámetros de ganancia, esta función de agrupamiento espacial produce un parámetro indicativo de la calidad más desfavorable en el espacio
mean (media)	Para cada índice temporal $t$ , calcular el promedio de todos los valores de parámetro. Esta función de agrupamiento espacial produce un parámetro que es indicativo de la calidad media en el espacio
std (desviación típica)	Para cada índice temporal $t$ calcular la desviación típica de todos los valores de parámetro. Esta función de agrupamiento espacial produce un parámetro que es indicativo de las variaciones de calidad en el espacio
below5%tail (resto por debajo del 5%)	Para cada índice temporal $t$ , ordenar los valores de parámetro de menor a mayor. Calcular el promedio de todos los valores de parámetro inferiores o iguales al nivel umbral del 5%, y a continuación restar el nivel del 5% de este promedio. Para los parámetros de atenuación, esta función de agrupamiento espacial permite medir el margen de niveles de calidad más desfavorables en el espacio. Es útil para medir los efectos sobre la calidad perceptual de las distorsiones localizadas espacialmente
above99%tail (resto por encima del 99%)	Para cada índice temporal $t$ , ordenar los valores de parámetro de menor a mayor. Calcular el promedio de todos los valores de parámetro mayores o iguales que el nivel umbral del 99%, y a continuación restar a este promedio el nivel del 99%. Para los parámetros de ganancia, esta función de agrupamiento espacial permite la medida del margen de los niveles de calidad más desfavorables en el espacio. Es útil para medir los efectos sobre la calidad perceptual de las distorsiones localizadas espacialmente

CUADRO 13

**Funciones de agrupamiento temporal y sus definiciones**

<b>Función de agrupamiento temporal</b>	<b>Definición</b>
10%	Ordenar la historia temporal de los valores de parámetro de menor a mayor y seleccionar el nivel de umbral del 10%. Para los parámetros de atenuación, esta función de agrupamiento temporal produce un parámetro indicativo de la calidad más desfavorable en el tiempo. Para los parámetros de ganancia, produce un parámetro indicativo de la mejor calidad en el tiempo
25%	Ordenar la historia temporal de los valores de parámetro de menor a mayor y seleccionar el nivel umbral del 25%
50%	Ordenar la historia temporal de los valores de parámetro de menor a mayor y seleccionar el nivel umbral del 50%
90%	Ordenar la historia temporal de los valores de parámetro de menor a mayor y seleccionar el nivel umbral del 90%. Para los parámetros de atenuación, esta función de agrupamiento temporal produce un parámetro indicativo de la calidad más favorable en el tiempo. Para los parámetros de ganancia, produce un parámetro indicativo de la calidad más desfavorable en el tiempo
mean (media)	Calcular la media de la historia temporal de los valores de parámetro. Esto produce un parámetro indicativo de la calidad media en el tiempo
std (desviación típica)	Calcular la desviación típica de la historia temporal de los valores de parámetro. Esta función de agrupamiento temporal produce un parámetro indicativo de las variaciones de calidad en el tiempo
above90%tail (resto por encima del 90%)	Ordenar la historia de los valores de parámetro de menor a mayor y calcular la media de todos los valores de parámetro que son mayores o iguales que el nivel umbral del 90%, y restar a continuación a esta media el nivel del 90%. Para los parámetros de ganancia, esta función de agrupamiento temporal permite la medición del margen de niveles de calidad más desfavorables en el tiempo. Es útil para medir los efectos sobre la calidad perceptual de las distorsiones localizadas temporalmente

**8.5 Recorte y ajuste no lineal**

El parámetro  $p$  totalmente positivo o totalmente negativo agrupado temporalmente, véase el § 8.4, puede ajustarse a escala en razón de las relaciones no lineales entre el valor del parámetro y la calidad percibida. Es preferible suprimir las posibles relaciones no lineales antes de construir los modelos de calidad de vídeo (véase el § 9), ya que se utilizará un algoritmo lineal de mínimos cuadrados para determinar los pesos óptimos de los parámetros. Las dos funciones no lineales de ajuste a escala que podrían aplicarse son la función raíz cuadrada, representada por  $\sqrt{x}$ , y la función cuadrado, representada por  $x^2$ . Si se aplica la función  $\sqrt{x}$  a un parámetro totalmente negativo, el parámetro se hará primero totalmente positivo (es decir se tomará su valor absoluto).

Por último, podría aplicarse una función de recorte representada por  $\text{clip}_T$ , siendo  $T$  el umbral de recorte, para reducir la sensibilidad del parámetro a pequeñas degradaciones. La función de recorte sustituye cualquier valor de parámetro entre el nivel de recorte y cero por el nivel de recorte, y a continuación resta el nivel de recorte de todos los valores resultantes del parámetro. Esto se representa matemáticamente del siguiente modo:

$$\text{clip}_T(p) = \begin{cases} \text{máx}(p, T) - T & \text{si } p \text{ es totalmente positivo} \\ \text{mín}(p, T) - T & \text{si } p \text{ es totalmente negativo} \end{cases}$$

## 8.6 Convenio de denominación de parámetros

En este apartado se resumen los convenios técnicos de denominación utilizados para los parámetros de calidad de vídeo. Este convenio asigna a cada parámetro un nombre largo consistente en palabras definitorias (subnombres) separadas por subrayados. El nombre técnico del parámetro resume el proceso exacto utilizado para su cálculo. Cada subnombre identifica una función o paso del proceso de cálculo del parámetro. Los subnombres se listan por su orden de aparición, de izquierda a derecha. El Cuadro 14 resume los subnombres utilizados para crear el nombre técnico del parámetro, listados por orden de aparición. El § 8.6.1 contiene ejemplos de nombres técnicos de parámetros y sus subnombres asociados del Cuadro 14.

CUADRO 14

### Convenio de denominación técnica utilizado para los parámetros de calidad de vídeo

Subnombre	Definición	Ejemplos
Color	Planos del espacio de color de la imagen utilizados por el parámetro	$Y$ para el plano de luminancia de la imagen. color para los planos de imagen ( $C_B, C_R$ )
Específico de la característica	El subnombre «Específico de la característica» describe los cálculos que hacen que este parámetro sea único. Todos los subnombres siguientes son procesos genéricos que pueden ser utilizados por muchos tipos distintos de parámetros. El subnombre «Específico de la característica» suele ser el nombre de la característica extraído del plano «color» en este punto del flujo, de aquí la posición de este subnombre. No obstante, también puede incluirse aquí información que no esté contemplada de otro modo en el convenio de denominación. Por ejemplo, el parámetro HV se aplica al subnombre «Estadística del bloque» por separado para los planos de imagen HV y $\overline{HV}$ . La relación subsiguiente de HV a $\overline{HV}$ se especifica mediante el nombre «Específico de la característica» (es decir en vez de ocupar un subnombre separado tras «Estadística del bloque»)	si13 para la característica $f_{SI13}$ del § 7.2.2. $hv13\_angleX.XXX\_rminYY$ para la característica $f_{HV13}$ del § 7.2.2, donde $X.XXX$ es $\Delta\theta$ e $YY$ es el $r_{min}$ . coher_color para la característica $f_{COHER\_COLOR}$ del § 7.3. cont para la característica $f_{CONT}$ de § 7.4. ati para la característica $f_{ATI}$ del § 7.5. contrast_ati para la característica $f_{CONTRAST\_ATI}$ del § 7.6
Desplazamiento del bloque	Presente cuando los bloques S-T se deslizan (por ejemplo solapándose en el tiempo). Cuando falta, se supone que los bloques son adyacentes en el tiempo	Deslizante

## CUADRO 14

## Convenio de denominación técnica utilizado para los parámetros de calidad de vídeo

Subnombre	Definición	Ejemplos
Imagen completa	Presente cuando el tamaño del bloque S-T contiene la región válida completa de la imagen. Cuando falta, debe estar presente el subnombre «tamaño del bloque»	Imagen
Tamaño del bloque	Presente cuando la imagen está dividida en bloques S-T (véase el § 7.1.1). Por coherencia, el tamaño del bloque siempre se indica relativo a las líneas de cuadro y píxeles de cuadro del plano de luminancia (Y). Por consiguiente, para vídeo muestreado 4:2:2, los bloques de color contendrán realmente la mitad del número especificado de píxeles horizontalmente. Cuando falta, debe estar presente el subnombre «imagen completa»	8 × 8 para bloques que tengan 8 líneas de cuadro verticalmente por 8 píxeles de cuadro horizontalmente. 128 × 128 para bloques que tengan 128 líneas de cuadro verticalmente por 128 píxeles de cuadro horizontalmente
Cuadros de bloque	Indica la amplitud temporal de los bloques S-T (véase el § 7.1.1), con referencia a vídeo de 30 fps. Por ejemplo, se utiliza 6F para representar un quinto de segundo, con independencia de la velocidad de cuadro del vídeo que se mide (por ejemplo 5 cuadros en un sistema de 25 fps, 3 cuadros en un sistema de 15 fps y 2 cuadros en un sistema de 10 fps)	1F para una amplitud temporal de un cuadro. 6F para una amplitud temporal de un quinto de segundo
Estadística de bloque	Función estadística utilizada para extraer la característica de cada región S-T, que genera un número para cada bloque S-T de píxeles. Presente salvo que «tamaño del bloque» = $I \times I$ (es decir 1 píxel). Antes de aplicar la estadística del bloque, los resultados intermedios contienen historias temporales de imágenes con un número por píxel (es decir imágenes filtradas); después, los resultados intermedios contienen un número por cada región S-T (es decir imágenes de la característica). Los parámetros que tengan dos planos de imagen (por ejemplo <i>hv13</i> y <i>coher_color</i> ) aplicarán la estadística de bloque por separado a ambos planos de imagen, generando dos imágenes de característica	mean es el promedio de los valores de los píxeles. std es la desviación típica de los valores de los píxeles. rms es el valor cuadrático medio de los valores de los píxeles
Umbral de perceptibilidad	Los valores producidos por la «estadística de bloque» pueden recortarse en un umbral de perceptibilidad <i>P</i> . Los valores entre cero y este umbral se sustituyen por el umbral	3 para un valor mínimo de característica de 3,0. 12 para un valor mínimo de característica de 12,0
Función de comparación	Función utilizada para comparar las características extraídas de las secuencias original y procesada de la característica (véase el § 8.2). Antes de la función de comparación, los resultados intermedios contienen historias temporales de las imágenes original y procesada de la característica; después, los resultados intermedios contienen una historia temporal de las imágenes del parámetro	log_gain (véase el § 8.2.1). ratio_loss (véase el § 8.2.1). euclid (véase el § 8.2.2)

CUADRO 14

## Convenio de denominación técnica utilizado para los parámetros de calidad de vídeo

Subnombre	Definición	Ejemplos
Función de agrupamiento espacial	Véase el § 8.3. Esta función se aplica a cada una de las imágenes de los parámetros (por ejemplo, todas las regiones S-T con el mismo índice temporal) y produce una historia temporal de los valores del parámetro. Antes del agrupamiento espacial, los resultados intermedios consisten en imágenes de parámetros conteniendo un valor para cada bloque S-T; después, los resultados intermedios son una historia temporal de números (es decir la historia temporal del parámetro). Debe estar presente para todos los parámetros excepto para los parámetros «Imagen completa»	Véase el Cuadro 12
Función de agrupamiento temporal	Véase el § 8.4. Esta función se aplica a la historia temporal del parámetro y produce un valor de parámetro para la totalidad de la secuencia de vídeo. Tras la agrupación temporal, el parámetro contiene valores que son todos negativos o todos positivos, pero no ambos. El cero se asocia a la ausencia de degradación, y los valores de parámetro tienen mayor degradación cuanto más se alejan de cero. Debe estar presente para todos los parámetros	Véase el Cuadro 13
Función no lineal	Véase el § 8.5. El examen de los valores del parámetro puede indicar que el parámetro debe ajustarse de manera no lineal para controlar linealmente el estado de los datos subjetivos. La función no lineal ejecuta este ajuste final. Si se aplica la función sqrt a un parámetro totalmente negativo, el parámetro se hará primero totalmente positivo (es decir se tomarán los valores absolutos)	sqrt para la raíz cuadrada del valor del parámetro temporalmente agrupado. square para el cuadrado del valor del parámetro agrupado temporalmente
Función de recorte	Véase el § 8.5. El examen final de los valores de parámetro puede indicar la necesidad de reducir aún más la sensibilidad del parámetro a pequeñas degradaciones (por ejemplo, valores de parámetro próximos a cero). Sustituir cualquier valor entre el nivel de recorte $T$ y cero por el nivel de recorte, y restar a continuación el nivel de recorte de todos los valores de parámetros resultantes	clip_0.45 Si los valores del parámetro son positivos, sustituir todos los valores menores que 0,45 por 0,45 y a continuación restar 0,45 de todos los valores del parámetro. Si los valores del parámetro son negativos, sustituir todos los valores mayores que -0,45 por -0,45 y a continuación añadir 0,45 a todos los valores del parámetro



### 8.6.1 Ejemplos de nombres de parámetros

En este apartado se incluyen cinco ejemplos de nombres técnicos, y una descripción paso a paso del procedimiento de subdenominación del Cuadro 14.

`Y_si13_8x8_6F_std_6_ratio_loss_below5%_mean`

*Y* significa que se utiliza el plano de imagen de luminancia. *si13* representa el filtrado de dichas imágenes con las  $13 \times 13$  máscaras espaciales del § 7.2.1 para preparar la extracción de la característica  $f_{SI13}$  de § 7.2.2.  $8 \times 8\_6F$  representa la división de la secuencia de vídeo en regiones S-T conteniendo ocho líneas de cuadro verticalmente por ocho píxeles horizontalmente por un quinto de segundo temporalmente (es decir 6 cuadros NTSC o 5 cuadros PAL). *std* representa la desviación típica de cada bloque. *6* representa la aplicación de un umbral de perceptibilidad, que sustituye todos los valores de desviación típica por debajo de 6,0 por un valor de 6,0. *ratio\_loss* representa la comparación entre las características original y procesada de cada bloque utilizando la función *ratio\_loss*. *below5%* representa el agrupamiento espacial de los valores del parámetro en cada índice temporal utilizando la función *below5%*. *mean* representa el agrupamiento temporal de la historia temporal del parámetro utilizando la función *mean*.

`color_coher_color_8x8_1F_mean_euclid_std_10%_clip_0.8`

*color* representa la utilización de los planos de imagen  $C_B$  y  $C_R$ . *coher\_color* representa el mantenimiento de la relación de fase entre las imágenes de  $C_B$  y  $C_R$  (tratándolas por separado) para preparar la extracción de la característica  $f_{COHER\_COLOR}$  del § 7.3.  $8 \times 8\_1F$  representa la división de cada cuadro en bloques de 8 líneas de cuadro de alto por 4 píxeles  $C_B$  y  $C_R$  de ancho (debido al submuestreo 4:2:2 de los planos de imagen  $C_B$  y  $C_R$ ) por 1 cuadro en el tiempo. *mean* representa el valor medio de cada bloque. *euclid* el cálculo de la distancia euclidiana entre los vectores originales ( $C_B$ ,  $C_R$ ) y los procesados ( $C_B$ ,  $C_R$ ) para cada bloque S-T. *std* representa la función de agrupamiento espacial *std*. *10%* representa la función de agrupamiento temporal *10%*. *clip\_0.8* representa el recorte del valor final del parámetro a un mínimo de 0,8 (es decir sustituyendo todos los valores entre 0,8 por 0,8, y restando a continuación 0,8).

`Y_hv13_angle0.225_rmin20_8x8_6F_mean_3_ratio_loss_below5%_mean_square_clip_0.05`

*Y* significa que se utiliza el plano de imagen de luminancia. *hv13* representa el filtrado de las imágenes *Y* con las  $13 \times 13$  máscaras espaciales del § 7.2.1 para preparar la extracción de la característica  $f_{HV13}$  del § 7.2.2 (es decir las imágenes HV y  $\overline{HV}$  se crean y tratan por separado hasta después del umbral de perceptibilidad). *angle0.225* y *rmin20* representan un  $\Delta\theta$  de 0,225 radianes y un  $r_{min}$  de 20 para el cálculo de la característica  $f_{HV13}$ .  $8 \times 8\_6F$  representa la división de la secuencia de vídeo en regiones S-T que contienen ocho líneas de cuadro verticalmente por ocho píxeles horizontalmente por un quinto de segundo temporalmente (es decir 6 cuadros NTSC o 5 cuadros PAL). *mean* representa el valor medio de cada bloque S-T para HV y  $\overline{HV}$ . *3* representa la aplicación de un umbral de perceptibilidad a dichas medias, sustituyendo todos los valores menores que 3 por 3,0. A continuación se calcula la característica  $f_{HV13}$  del § 7.2.2 como relación de las medias recortadas de HV a las medias recortadas de  $\overline{HV}$ , como se especifica en *hv13\_angle0.225\_rmin20*. El subnombre de característica específica *ratio\_loss* representa la utilización de la función de comparación *ratio\_loss* para cada característica  $f_{HV13}$  original y su correspondiente procesada extraídas de un bloque S-T. *below5%* especifica la función de agrupamiento espacial. *mean* especifica la función de agrupamiento temporal. *square* especifica la función no

lineal para cada valor de parámetro agrupado en el tiempo. clip\_0.05 representa la función de recorte, en la que se sustituyen todos los valores inferiores a 0,05 por 0,05, restándose a continuación 0,05 del resultado (recuérdese que el parámetro totalmente negativo se convertirá en totalmente positivo debido a la función no lineal square).

Y\_contrast\_ati\_4x4\_6F\_std\_3\_ratio\_gain\_mean\_10%

Y significa que se utiliza el plano de luminancia. contrast\_ati representa el cálculo de dos versiones filtradas separadas de la imagen para preparar la extracción de la característica  $f_{CONTRAST\_ATI}$  del § 7.6. El primer filtro, contrast, considerará los planos de luminancia directamente (véase el § 7.4). El segundo filtro, ati, considerará imágenes generadas tomando las diferencias entre planos de luminancia correlativos (véase el § 7.5). Las imágenes contrast y ati se tratan por separado hasta después de la fijación de umbrales. 4x4\_6F significa que las dos secuencias de vídeo se dividen en regiones S-T que contienen cuatro líneas de cuadro verticalmente por cuatro píxeles horizontalmente por un quinto de segundo temporalmente (por ejemplo 6 cuadros NTSC o 5 cuadros PAL). El primer bloque S-T de imágenes ati sólo contendrá realmente 5 imágenes en vez de 6 ya que no es posible generar una imagen ati para el primer cuadro de la secuencia (es decir no hay disponible ninguna imagen anterior en el tiempo). Esta excepción se especifica como parte del subnombre específico de la característica. std representa la desviación típica de cada bloque. A continuación, como se especifica en el subnombre específico de la característica del § 7.6, se aplica un umbral de perceptibilidad de 3 a las características contrast y ati (sustituyendo todos los valores menores que 3 por 3,0). Después, se multiplica el valor de bloque contrast por el valor de bloque ati para cada bloque S-T (véase la Nota de pie de página del § 7.6 sobre instrucciones especiales para efectuar esta multiplicación) y se continúan los cálculos con esta imagen combinada de la característica. ratio\_gain es la función de comparación utilizada para comparar cada característica original y procesada de los bloques S-T. mean es la función de agrupamiento espacial. 10% es la función de agrupamiento temporal.

## 9 Modelo general

En este apartado se ofrece una descripción completa del modelo general VQM (representada por  $VQM_G$ ). El modelo general se optimiza para alcanzar una correlación máxima objetiva a subjetiva utilizando una amplia gama de calidades de vídeo y velocidades binarias. El modelo general tiene parámetros objetivos para medir los efectos perceptuales de una amplia gama de degradaciones tales como la borrosidad, la distorsión de bloque, el movimiento entrecortado/innatural, el ruido (tanto en el canal de luminancia como en el de crominancia), y los bloques con errores (por ejemplo, lo que podría verse normalmente cuando hay errores en la transmisión digital). Este modelo consta de una combinación lineal de parámetros de calidad de vídeo cuyos convenios de denominación se describen en el § 8.6. La selección de los parámetros de calidad de vídeo se determinó por los criterios de optimización citados. El modelo general produce valores de salida que van de cero (no hay degradación percibida) a uno (máxima degradación percibida) aproximadamente. Para colocar los resultados en la escala de calidad continua de doble estímulo (DSCQS), hay que multiplicar  $VQM_G$  por 100.

El modelo general se diseñó basándose en vídeo de la Recomendación UIT-R BT.601 evaluado subjetivamente a una distancia de observación de seis alturas de imagen. Cuando se analizan secuencias de vídeo con diferentes distancias de observación, debe aplicarse un factor de ajuste a los resultados. Conforme aumenta la distancia de observación, las degradaciones resultan menos visibles; al disminuir la distancia de observación, las degradaciones resultan más visibles. Debe realizarse con cuidado La comparación de resultados de secuencias de vídeo observadas desde diferentes distancias debe llevarse a cabo con la necesaria precaución.

$VQM_G$  consta de una combinación lineal de siete parámetros. Cuatro de ellos se basan en características extraídas de los gradientes espaciales de la componente de luminancia  $Y$  (véase el § 7.2.2), dos se basan en características extraídas del vector formado por las dos componentes de crominancia ( $C_B$ ,  $C_R$ ) (véase el § 7.3), y uno se basa en las características del contraste y la información temporal absoluta, extraídos de la componente de luminancia  $Y$  (véanse los § 7.4 y 7.5, respectivamente).  $VQM_G$  viene dado por:

$$\begin{aligned}
 VQM_G = & \{-0.2097 * Y_{si13\_8\times8\_6F\_std\_12\_ratio\_loss\_below5\%\_10\%} \\
 & +0.5969 * Y_{hv13\_angle0.225\_rmin20\_8\times8\_6F\_mean\_3\_ratio\_loss\_below5\%\_mean\_square\_clip\_0.06} \\
 & +0.2483 * Y_{hv13\_angle0.225\_rmin20\_8\times8\_6F\_mean\_3\_log\_gain\_above95\%\_mean} \\
 & +0.0192 * color\_coher\_color\_8\times8\_1F\_mean\_euclid\_std\_10\%\_clip\_0.6} \\
 & -2.3416 * [Y_{si13\_8\times8\_6F\_std\_8\_log\_gain\_mean\_mean\_clip\_0.004} |^{0.14}] \\
 & +0.0431 * Y_{contrast\_ati\_4\times4\_6F\_std\_3\_ratio\_gain\_mean\_10\%} \\
 & +0.0076 * color\_coher\_color\_8\times8\_1F\_mean\_euclid\_above99\%tail\_std} |_{0.0}
 \end{aligned}$$

Recuérdese que las características anteriores para el modelo general con una amplitud de tiempo «6F» contendrán realmente cinco cuadros de vídeo PAL (625 líneas).

El cuadrado del parámetro  $hv\_loss$  es necesario para que la respuesta del parámetro pueda hacerse lineal con respecto a los datos subjetivos. Obsérvese que puesto que el parámetro  $hv\_loss$  se convierte en positivo tras el cuadrado, se utiliza un factor de ponderación positivo. Obsérvese asimismo que el parámetro  $hv\_loss$  se recorta a 0,06, el parámetro de color se recorta a 0,6, y el parámetro  $si\_gain$  se recorta a 0,004. El parámetro  $si\_gain$  es el único parámetro de mejora de calidad del modelo (ya que el parámetro  $si\_gain$  es positivo, resulta un factor de ponderación negativo en las contribuciones negativas a  $VQM$  que producen mejoras de calidad). El parámetro  $si\_gain$  mide las mejoras de calidad resultantes del afinado o realce de los bordes. El recorte del parámetro en un umbral superior de 0,14 justo antes de multiplicarlo por el factor de ponderación del parámetro evita mejoras excesivas de la  $VQM$  superiores a 1/3, aproximadamente, de una unidad de calidad, que es la mejora máxima observada en el conjunto general de datos subjetivos (es decir un HRC sólo se recompensará por un pequeño realce de los bordes).

La  $VQM$  total (una vez sumadas las contribuciones de todos los parámetros) se recorta en un umbral inferior de 0,0 para evitar números  $VQM$  negativos. Por último, se aplica a los valores de  $VQM$  superiores a 1,0 una función de aplastamiento que permite una sobreoscilación máxima del 50%, para limitar los valores de  $VQM$  de vídeos excesivamente distorsionados que queden fuera del rango de datos subjetivos actualmente disponibles.

$$\text{Si } VQM_G > 1,0, \text{ entonces } VQM_G = (1 + c) * VQM_G / (c + VQM_G), \text{ siendo } c = 0,5.$$

Si se calcula  $VQM_G$  del modo anterior, se obtendrán valores mayores o iguales que cero y un valor máximo nominal de uno. Ocasionalmente,  $VQM_G$  puede ser mayor que uno en las escenas de vídeo extremadamente distorsionadas.

## 10 Referencias Bibliográficas

- JAIN, A. K. [1989] *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall Inc., p. 348-357.
- Norma SMPTE [1995a] SMPTE 125M. Television – Component Video Signal 4:2:2 – Bit-Parallel Digital Interface. Society of Motion Picture and Television Engineers, 595 West Hartsdale Avenue, White Plains, NY 10607.
- Norma SMPTE [1995b] SMPTE Recommended Practice 187. Center, Aspect Ratio, and Blanking of Video Images. Society of Motion Picture and Television Engineers, 595 West Hartsdale Avenue, White Plains, NY 10607.
- Norma SMPTE [1999] SMPTE 170M. Television composite Analog Video Signal – NTSC for Studio Applications. Society of Motion Picture and Television Engineers, 595 West Hartsdale Avenue, White Plains, NY 10607.
- PINSON, M. y WOLF, S. [febrero de 2002] Video Quality Measurement User's Manual. NTIA Hybook 02-1. National Telecommunications and Information Administration.
- WOLF, S. y PINSON M. [12-13 de noviembre de 1998] In-service performance metrics for MPEG-2 video systems. Proc. Made to Measure 98 – Measurement Techniques of the Digital Age Technical Seminar, Conferencia Técnica patrocinada conjuntamente por la International Academy of Broadcasting (IAB), la UIT y la Technical University of Braunschweig (TUB), Montreux, Suiza.
- WOLF, S. y PINSON, M. [septiembre de 1999] Spatial-temporal distortion metrics for in-service quality monitoring of any digital video system. Proc. SPIE International Symposium on Voice, Video, and Data Communications, Boston, MA.
- WOLF, S. y PINSON, M. [julio de 2001] The relationship between performance and spatial-temporal region size for reduced-reference, in-service video quality monitoring systems. Proc. SCI/ISAS 2001 (Systematics, Cybernetics, and Informatics/Information Systems Analysis and Synthesis), p. 323-328.
- WOLF, S. y PINSON, M. [junio de 2002] Video Quality Measurement Techniques. NTIA Report 02-392. National Telecommunications and Information Administration.

## Anexo 5a

### Datos objetivos brutos de la VQM de la NTIA

El presente Anexo tiene por objeto divulgar la totalidad de los datos objetivos brutos de la VQM de la NTIA.

## Resumen de datos brutos

Este modelo general desarrollado por la NTIA se diseñó originalmente para obtener valores en una escala nominal de 0 a 1, representando el 0 la ausencia de percepción de degradación y el 1 el máximo de degradación percibida. No obstante, el ejecutable en binario sometido a la prueba VQEG Fase II FR-TV transformó los valores (0, 1) del modelo general a (0, 100) para hacerlos compatibles con la DSCQS. Como todos los valores del modelo deben ajustarse a (0, 1), se ha suprimido el factor multiplicativo de 100 (es decir la multiplicación por 100) para restaurar la escala original (0, 1) del modelo general.

Los valores del modelo general se han calculado utilizando los 8 s centrales de vídeo de cada corto, descartando los 10 cuadros suplementarios de vídeo al principio y al final de cada fichero de vídeo descrito en el plan de pruebas VQEG Fase II FR-TV. Para las rutinas de calibración se utilizó una incertidumbre de 30 cuadros y una frecuencia de 15 cuadros (véase el § 6 del Anexo 5). Además, la SROI utilizada para calcular el valor de la VQM de cada corto se escogió del siguiente modo:

*Paso 1:* Para los sistemas de vídeo de 525 líneas, se utilizó una SROI por defecto de 672 píxeles  $\times$  448 líneas centrada en el cuadro de vídeo. Para los sistemas de 625 líneas, se utilizó una SROI por defecto de 672 píxeles  $\times$  544 líneas centrada en el cuadro de vídeo. Estas SROI por defecto pueden modificarse como se indica en los Pasos 2 y 3.

*Paso 2:* El modelo requiere 6 píxeles/líneas adicionales válidos en todos los lados de la citada SROI para que los filtros espaciales funcionen correctamente. Si la PVR (calculada automáticamente con arreglo al § 6.2 del Anexo 5) no es bastante grande para abarcar la SROI por defecto más 6 píxeles/líneas (Paso 1), se reduce la SROI en múltiplos de 8 píxeles/líneas únicamente en el sentido necesario (horizontal o vertical).

*Paso 3:* La SROI siempre se centra horizontalmente de modo que la muestra izquierda comience en un punto de muestreo de luminancia/crominancia de la Recomendación UIT-R BT.601 que se encuentre en la misma posición. La SROI se centra verticalmente de modo que cuando se divide en dos campos, se descarta el mismo número de líneas de la parte superior de cada campo. Si se ha reducido el tamaño de la SROI en el Paso 2, tal vez no sea posible centrar perfectamente la SROI en el cuadro de vídeo.

El soporte lógico de evaluación que implementa el modelo general y sus rutinas de calibración puede descargarse de:

<http://www.its.bldrdoc.gov/n3/video/vqmssoftware.htm>

## CUADRO 15

## Datos objetivos brutos para 525 líneas

N.º Origen	N.º HRC	NTIA: Proponente H			
1	1	0,660 <sup>(1)</sup>	9	14	0,124
1	2	0,347	10	9	0,666
1	3	0,286	10	10	0,250
1	4	0,178	10	11	0,375
2	1	0,449	10	12	0,129
2	2	0,246	10	13	0,078
2	3	0,119	10	14	0,153
2	4	0,061	11	9	0,513
3	1	0,321	11	10	0,534
3	2	0,167	11	11	0,407
3	3	0,076	11	12	0,161
3	4	0,049	11	13	0,148
4	5	0,396	11	14	0,159
4	6	0,280	12	9	0,600
4	7	0,222	12	10	0,410
4	8	0,183	12	11	0,471
5	5	0,329	12	12	0,244
5	6	0,217	12	13	0,171
5	7	0,159	12	14	0,114
5	8	0,115	13	9	0,537
6	5	0,542	13	10	0,425
6	6	0,266	13	11	0,346
6	7	0,189	13	12	0,215
6	8	0,139	13	13	0,188
7	5	0,258	13	14	0,169
7	6	0,161			
7	7	0,108			
7	8	0,076			
8	9	0,911			
8	10	0,717			
8	11	0,721			
8	12	0,526			
8	13	0,424			
8	14	0,311			
9	9	0,827			
9	10	0,453			
9	11	0,512			
9	12	0,264			
9	13	0,188			

<sup>(1)</sup> Para la fuente 1, HRC 1, el soporte lógico de calibración entregado al VQEG provocó un error de registro espacial/temporal que produjo la estimación incorrecta del vídeo procesado a reencuadrar (es decir lo desplazó en un campo, véase el § 6.1.2 en el Anexo 5). Para las otras escenas de HRC 1, el registro espacial/temporal se estimó correctamente. El § 6.1.5.7 del Anexo 5, recomienda el filtrado de mediana de los resultados de calibración en todas las escenas de un determinado HRC como método para producir estimaciones de calibración de mayor validez para un determinado HRC. No obstante, el plan de la prueba VQEG Fase II especificaba que todo el soporte lógico VQM debía producir una única estimación de calidad para cada corto por separado. Así pues el filtrado de mediana de los números de calibración en todas las escenas para un determinado HRC no estaba autorizado por el plan de prueba. De haberse autorizado el filtrado de mediana de los números de calibración, el soporte lógico VQM habría registrado correctamente este corto y la puntuación objetiva sin procesar habría sido 0,529.

CUADRO 16

**Datos objetivos brutos para 625 líneas**

<b>N.º de fuente</b>	<b>N.º de HRC</b>	<b>NTIA: proponente H</b>			
1	2	0,421	7	10	0,270
1	3	0,431	8	4	0,345
1	4	0,264	8	6	0,311
1	6	0,205	8	9	0,280
1	8	0,155	8	10	0,242
1	10	0,123	9	4	0,344
2	2	0,449	9	6	0,285
2	3	0,473	9	9	0,246
2	4	0,312	9	10	0,192
2	6	0,260	10	4	0,410
2	8	0,226	10	6	0,355
2	10	0,145	10	9	0,313
3	2	0,472	10	10	0,241
3	3	0,506	11	1	0,739
3	4	0,308	11	5	0,468
3	6	0,239	11	7	0,199
3	8	0,183	11	10	0,201
3	10	0,146	12	1	0,548
4	2	0,409	12	5	0,441
4	3	0,458	12	7	0,367
4	4	0,384	12	10	0,307
4	6	0,354	13	1	0,598
4	8	0,280	13	5	0,409
4	10	0,232	13	7	0,321
5	2	0,470	13	10	0,277
5	3	0,521			
5	4	0,260			
5	6	0,234			
5	8	0,132			
5	10	0,083			
6	2	0,391			
6	3	0,364			
6	4	0,290			
6	6	0,252			
6	8	0,181			
6	10	0,169			
7	4	0,422			
7	6	0,385			
7	9	0,336			

## Apéndice 1

### Resultados de la prueba del Grupo de Expertos en calidad de vídeo FR-TV Fase II

#### 1 Introducción

La calidad de funcionamiento de los modelos de calidad perceptual contemplados en esta Recomendación se valoró mediante dos evaluaciones paralelas. En la primera evaluación, se utilizó un método subjetivo estándar, el método DSCQS, para obtener los índices de calidad subjetivos del material de vídeo por grupos de observadores expertos. En la segunda evaluación, los índices objetivos se obtuvieron mediante modelos computacionales objetivos. Para cada modelo, se computaron varias métricas a fin de medir la precisión y coherencia con la que los índices objetivos predecían los subjetivos.

En este Apéndice se describe la porción de evaluación subjetiva de la prueba así como los resultados de los modelos computacionales objetivos presentados por los siguientes proponentes:

- Modelo 1 (British Telecom; identificado en VQEG FR-TV Fase II como proponente D).
- Modelo 2 (Yonsei University/Radio Research Laboratory/SK Telecom; identificado en VQEG FR-TV Fase II como proponente E).
- Modelo 3 (CPqD; identificado en VQEG FR-TV Fase II como proponente F).
- Modelo 4 (NTIA; identificado en VQEG FR-TV Fase II como proponente H).

Las pruebas subjetivas se llevaron a cabo en tres laboratorios independientes. En dos de ellos Communications Research Center (CRC, Canadá) y Verizon (Estados Unidos de América), se realizó la prueba con secuencias de 525/60 Hz mientras que en un tercer laboratorio, Fondazione Ugo Bordoni (FUB, Italy), se realizó la prueba con secuencias de 625/50 Hz.

En el Documento<sup>1</sup> mencionado se presenta una descripción detallada de la prueba del Grupo de Expertos en Calidad de Vídeo FR-TV Fase II.

#### 2 Materiales de vídeo

Las secuencias de vídeo de prueba en los formatos de 525/60 y 625/50 líneas se ajustaban al formato de vídeo de componentes 4:2:2 de la Recomendación UIT-R BT.601 con una relación de formato de 4:3.

##### 2.1 SRC y HRC

Para cada una de las pruebas con 525 y 625 líneas, se utilizaron 13 secuencias fuente (SRC) de distintas características (por ejemplo formato, información temporal y espacial, color, etc.) (véanse los Cuadros 17 y 18).

En ambas pruebas, se escogieron HRC representativos de condiciones típicas de la distribución secundaria de vídeo con calidad de televisión codificado digitalmente. En la prueba de 625 líneas se utilizaron 10 HRC; sus características se recogen en el Cuadro 19. En la de 525 líneas se utilizaron 14 HRC, sus características se recogen en el Cuadro 20.

Tanto en las pruebas de 625 líneas como en las de 525 líneas, las SRC y los HRC se combinaron en una matriz dispersa (véanse los Cuadros 23-26).



## CUADRO 17

## Secuencias del formato 625/50 (SRC)

Número del SRC	Características
1	Vista del horizonte desde un barco en movimiento; producida en película de 16:9, pasada a telecine de 576i/50
2	Bailarines en suelo de madera moviéndose con rapidez, con escaso detalle; originalmente capturado en formato D5
3	Partido de balonvolea masculino en pista cubierta; capturado en formato D5
4	Fútbol femenino en acción con enfoque rápido de cámara; capturado en D5
5	Dibujos animados tradicionales de 12 fps; fuente pasada a película de 24 fps, pasada después a telecine de 576i/50
6	Meridianos y paralelos del globo terráqueo en rotación lenta; capturada en DigiBetaCam
7	Movimiento rápido de la escena y de la cámara, con efectos de iluminación
8	Tocando la guitarra en primer plano, con efectos de luz cambiantes
9	Color, movimiento, detalle
10	Gran detalle, fondo con textura, movimiento
11	Color, movimiento, detalle
12	Partido de rugby en exteriores; movimiento, color
13	Movimiento, detalles, agua en movimiento
14 (demo)	Movimiento rápido de la escena y de la cámara, con efectos de iluminación
15 (demo)	Movimiento rápido de la escena y de la cámara, con efectos de iluminación
16 (demo)	Primer plano de un rostro seguido de toma panorámica de una obra

CUADRO 18  
Secuencias del formato 525/60 (SRC)

Número del SRC	Características
1	Partido de fútbol en exteriores, con color, movimiento y fondo basto
2	Paisaje otoñal con color detallado, zum lento
3	Dibujos animados con movimiento, color y cortes de escena
4	Escena de parque con gran detalle, con agua; procedente de una fuente de TVAD
5	Color y movimiento rápido; procedente de una fuente de TVAD
6	Color, gran superficie de agua; procedente de una fuente de TVAD
7	Partido de fútbol de aficionados, movimiento moderado; procedente de TVAD
8	Parque acuático; (DigiBetaCam)
9	Actividad en parque temático con movimiento moderado, gran detalle, zum lento; (DigiBetaCam)
10	Color, movimiento, iluminación moderadamente baja; (DigiBetaCam)
11	Dibujos animados tradicionales a 12 fps, pasado a película de 24 fps y pasada a telecine de 480i/60
12	Detalle de fuente en exteriores con zum de cámara; (DigiBetaCam)
13	Cambios de escena, de primeros planos de arranque de motor a tomas panorámicas lejanas, y viceversa; película original pasada a telecine de 480i/60
14 (demo)	Primer plano de rosa con brisa suave; movimiento, color y detalle; (DigiBetaCam)
15 (demo)	Gran detalle, escaso movimiento, procedente de TVAD
16 (demo)	Estatuas en rotación lenta, ramas de árbol en movimiento; (DigiBetaCam)

CUADRO 19  
HRC 625/50

Número del HRC	Velocidad binaria	Resolución	Método	Observaciones
1	768 kbit/s	CIF	H.263	Pantalla completa (HRC15 de VQEG 1)
2	1 Mbits/s	320H	MPEG2	Proponente codificado
3	1,5 Mbit/s	720H	MPEG2	Codificado por FUB
4	2,5 → 4 Mbit/s	720H	MPEG2	Puesto en cascada por FUB
5	2 Mbit/s	3/4	MPEG2 <a href="#">sp@ml</a>	HRC13 de VQEG 1
6	2,5 Mbit/s	720H	MPEG2	Codificado por FUB
7	3 Mbit/s	total	MPEG2	HRC9 de VQEG 1
8	3 Mbit/s	704H	MPEG2	Proponente codificado
9	3 Mbit/s	720H	MPEG2	Codificado por FUB
10	4 Mbit/s	720H	MPEG2	Codificado por FUB

## CUADRO 20

## HRC 525/60

Número del HRC	Velocidad binaria	Resolución	Método	Observaciones
1	768 kbit/s	CIF	H.263	Pantalla completa (HRC15 de VQEG 1)
2	2 Mbit/s	3/4	MPEG2, <a href="#">sp@ml</a>	HRC13 de VQEG 1
3	3 Mbit/s	total	MPEG2	HRC9 de VQEG 1
4	5 Mbit/s	720H	MPEG2	Codificado por CRC
5	2 Mbit/s	704H	MPEG2	Codificado por CRC
6	3 Mbit/s	704H	MPEG2	Codificado por CRC
7	4 Mbit/s	704H	MPEG2	Codificado por CRC
8	5 Mbit/s	704H	MPEG2	Codificado por CRC
9	1 Mbit/s	704H	MPEG2	Proponente codificado; baja velocidad binaria combinada con alta resolución
10	1 Mbit/s	480H	MPEG2	Codificado por CRC; baja velocidad binaria, baja resolución
11	1,5 Mbit/s	528H	MPEG2	Proponente codificado; modulación MAQ-64; salida NTSC compuesta convertida en componentes
12	4->2 Mbit/s	720H	MPEG2	Proponente codificado; codificadores en cascada
13	2,5 Mbit/s	720H	MPEG2	Codificado por CRC
14	4 Mbit/s	720H	MPEG2	Proponente codificado; utiliza códec de soporte lógico

### 3 Metodología para la evaluación de la calidad de funcionamiento del modelo objetivo

Para las pruebas subjetivas se utilizó el método DSCQS de la Recomendación UIT-R BT.500. Para la prueba de 525 líneas, se recogieron las notas medias de opinión de diferencias (DMOS, *difference mean opinion scores*) para 63 combinaciones SRC × HRC. Para la de 625 líneas, se recogieron las DMOS para 64 combinaciones SRC × HRC. Se obtuvieron asimismo los datos objetivos de estas mismas combinaciones SRC × HRC para cada modelo computacional objetivo.

Para la evaluación del modelo, se ajustaron los datos subjetivos y se transformaron los datos objetivos sin linealidad de acuerdo con una escala variable entre 0 (indistinguible de la fuente) y 1. La transformación no lineal fue la siguiente:

$$DMOS_p = b1 / (1 + \exp(-b2 * (VQR - b3)))$$

siendo:

$VQR$ : (índice de calidad de vídeo) valor real de salida del modelo computacional objetivo

$DMOS_p$ : valor transformado sin linealidad.

La calidad de funcionamiento de los modelos objetivos se evaluó en cuanto a tres aspectos de su capacidad de estimación de evaluación subjetiva de la calidad de vídeo:

- precisión de la predicción – capacidad de predecir los índices de calidad subjetivos con poco error;
- monotonía de la predicción – grado de concordancia de las predicciones del modelo con las magnitudes relativas de los índices de calidad subjetivos; y
- coherencia de la predicción – grado en el que se mantiene la precisión de la predicción del modelo en toda la gama de secuencias de prueba de vídeo, o sea robustez de su respuesta en relación con una diversidad de degradaciones de vídeo.

Estos atributos se evaluaron mediante siete métricas de calidad de funcionamiento descritas a continuación.

*Métrica 1:* Coeficiente de correlación lineal de Pearson entre  $DMOS_p$  y  $DMOS$ .

*Métrica 2:* Coeficiente de correlación del orden del rango de Spearman entre  $DMOS_p$  y  $DMOS$ .

La correlación de Spearman y la correlación de Pearson así como las demás estadísticas se calcularon para todas las combinaciones de SRC por HRC simultáneamente.

*Métrica 3:* Relación de valores atípicos entre los «puntos atípicos» y el total de puntos  $N$ .

$$\text{Relación de valores atípicos} = (\text{número total de valores atípicos})/N$$

donde un valor atípico es un punto para el que:  $ABS[ Q_{\text{error}}[i] ] > 2 * DMOS_{\text{StandardError}}[i]$ .

Se utilizó dos veces el error estándar de la DMOS como umbral para definir un punto atípico.

*Métricas 4, 5 y 6:* Estas métricas se evaluaron en base al método descrito en el Informe T1.TR.72-2001 [ATIS, 2001].

4. Error eficaz;

5. Potencia resolutive; y

6. Errores de clasificación.

Obsérvese que en la evaluación de los modelos con arreglo a este método se ha omitido el procedimiento de calibración cruzada descrito en el mismo, ya que no es pertinente para las mediciones de la calidad de funcionamiento de los modelos individuales.

*Métrica 7:* Esta métrica se basa en la prueba  $F$ . Se efectuaron dos mediciones de la prueba  $F$ . En la primera medición de  $F$  se utilizó el error cuadrático medio calculado a partir de los índices subjetivos individuales. El MSE se obtuvo para un modelo «nulo u óptimo», correspondiente a la DMOS observada y residuos asociados, y para cada uno de los modelos objetivos. Las pruebas  $F$  se llevaron a cabo para comparar el MSE del modelo nulo con el de cada modelo, y el MSE del modelo de mejor calidad de funcionamiento con el de los demás modelos. La segunda medición de  $F$  se basó en el MSE calculado a partir de los índices medios, es decir DMOS. Concretamente, se calcularon los MSE para cada modelo utilizando los residuos entre los DMOS predichos y observados. Se efectuaron pruebas  $F$  para comparar el MSE del modelo de mejor calidad de funcionamiento con el de los demás modelos.

#### 4 Evaluación de los resultados

Los resultados de los cálculos de las métricas se presentan en los Cuadros 21 y 22, uno correspondiente a los datos de 525 líneas y el otro a los de 625 líneas.

La concordancia de las siete métricas de los cuadros es prácticamente perfecta. Si un modelo objetivo se comporta bien con arreglo a una métrica suele hacerlo igualmente con las demás métricas, y viceversa. Además, la clasificación de los modelos objetivos en base a las distintas métricas es esencialmente idéntica en cada uno de los dos formatos de vídeo. No obstante, los resultados de las dos pruebas (525 y 625) son semejantes aunque no idénticos. Hay algunos cambios de clasificación ostensibles entre un experimento y el otro.

Los datos subjetivos ajustados a escala utilizados para calcular estas métricas se muestran en los Cuadros 23-26. Los correspondientes datos objetivos obtenidos de los cuatro modelos computacionales objetivos se presentan en los Anexos 2-5.

## 5 Datos de la PSNR

La PSNR es una métrica sencilla de la calidad de vídeo. La calidad de funcionamiento de las VQM puede compararse a la calidad de funcionamiento de la PSNR. La PSNR para las secuencias de prueba fue calculada por varios proponentes. Las métricas correspondientes a las PSNR más elevadas se presentan en los Cuadros 21 y 22.

CUADRO 21  
Resumen de los análisis para 525 líneas

Número de línea	Métrica	D525	E525	F525	H525	PSNR525
1	1. Correlación Pearson	0,937	0,857	0,835	0,938	0,804
2	2. Correlación Spearman	0,934	0,875	0,814	0,936	0,811
3	3. Relación de valores atípicos	33/63 = 0,52	44/63 = 0,70	44/63 = 0,70	29/63 = 0,46	46/63 = 0,73
4	4. Error eficaz, 63 puntos de datos	0,075	0,11	0,117	0,074	0,127
5	5. Potencia de resolución, delta VQM (tanto mejor cuanto menor)	0,2177	0,2718	0,3074	0,2087	0,3125
6	6. Porcentaje de errores de clasificación (mínimo sobre delta VQM)	0,1889	0,2893	0,3113	0,1848	0,3180
7	7. MSE del modelo/MSE del modelo óptimo	1,262	1,59	1,68	1,256	1,795
8	F = MSE del modelo/MSE del proponente H	1,005	1,266	1,338	1	1,429
9	MSE del modelo, 4219 puntos de datos	0,02421	0,03049	0,03223	0,02409	0,03442
10	MSE del modelo óptimo, 4219 puntos de datos	0,01918	0,01918	0,01918	0,01918	0,01918
11	MSE del modelo, 63 puntos de datos	0,00559	0,01212	0,01365	0,00548	0,01619
12	F= MSE63 modelo/MSE63 proponente H	1,02	2,212	2,491	1	2,954

NOTA 1 – Las Métricas 5 y 6 se calcularon con el código Matlab® publicado en T1.TR.72-2001.

NOTA 2 – La Métrica 5 se calculó visualmente de los diagramas de dispersión de los documentos de salida.

NOTA 3 – Los valores de la Métrica 7 inferiores a 1,07 indican que el modelo no es distinto, en cuanto a fiabilidad, del modelo óptimo.

NOTA 4 – Los valores de la línea 8 superiores a 1,07 indican que el modelo tiene residuos sensiblemente mayores que el modelo proponente mayor, *H* en este caso.

NOTA 5 – Los valores de la línea 12 superiores a 1,81 indican que el modelo tiene residuos sensiblemente mayores que el mayor modelo proponente, *H* en este caso.

CUADRO 22

## Resumen de los análisis para 625 líneas

Número de línea	Métrica	D625	E625	F625	H625	PSNR625
1	1. Correlación Pearson	0,779	0,87	0,898	0,886	0,733
2	2. Correlación Spearman	0,758	0,866	0,883	0,879	0,74
3	3. Relación de valores atípicos	28/64 = 0,44	24/64 = 0,38	21/64 = 0,33	20/64 = 0,31	30/64 = 0,47
4	4. Error eficaz, 64 puntos de datos	0,113	0,089	0,079	0,083	0,122
5	5. Potencia de resolución, delta VQM (tanto mejor cuanto menor)	0,321	0,281	0,270	0,267	0,313
6	6. Porcentaje de errores de clasificación (mínimo sobre delta VQM)	0,305	0,232	0,204	0,199	0,342
7	7. MSE del modelo /MSE del modelo nulo	1,652	1,39	1,303	1,339	1,773
8	F = MSE del modelo/MSE del proponente H	1,268	1,067	1	1,028	1,361
9	MSE del modelo, 1728 puntos de datos	0,02953	0,02484	0,02328	0,02393	0,03168
10	MSE del modelo nulo, 1728 puntos de datos	0,01787	0,01787	0,01787	0,01787	0,01787
11	MSE del modelo, 64 puntos de datos	0,0127	0,00786	0,00625	0,00693	0,01493
12	F= MSE64 del modelo/MSE64 del proponente F	2,032	1,258	1	1,109	2,389

NOTA 1 – Las Métricas 5 y 6 se calcularon con el código Matlab® publicado en T1.TR.72-2001.

NOTA 2 – La Métrica 5 se calculó a ojo a partir de los diagramas de dispersión de los documentos de salida.

NOTA 3 – Los valores de la Métrica 7 inferiores a 1,12 indican que el modelo no es distinto, en cuanto a fiabilidad, del modelo óptimo.

NOTA 4 – Los valores de la línea 8 superiores a 1,12 indican que el modelo tiene residuos sensiblemente mayores que el modelo proponente mayor, *F* en este caso.

NOTA 5 – En el caso de 625 líneas con 1728 observaciones, el valor crítico de la estadística de *F* es 1,12.

NOTA 6 – Los valores de la línea 12 superiores a 1,81 indican que el modelo tiene residuos sensiblemente mayores que el mayor modelo proponente, *F* en este caso.

CUADRO 23

Datos subjetivos para todas las combinaciones HRC-SRC 525/60 – (valores de DMOS)

SRC (Imagen)	HRC													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0,5402368	0,5483205	0,4024097	0,3063528										
2	0,5025558	0,3113346	0,1881739	0,1907347										
3	0,4682724	0,3088831	0,1300389	0,1293293										
4					0,6742005	0,4250873	0,3762656	0,2972294						
5					0,4682559	0,3203024	0,2071702	0,1652752						
6					0,5690291*	0,4370961	0,3591788	0,2482169						
7					0,3796362	0,2276934	0,1644409	0,1819566						
8									0,9513387	0,789748	0,8405916	0,5221555	0,4572049	0,4614104
9									0,8262912	0,660339	0,7100111	0,4921708	0,3656559	0,2960957
10									0,9084171	0,5908784	0,7302376	0,3345703	0,2565459	0,2953144
11									0,6675853	0,7054929	0,5761193	0,32761	0,310495	0,331051
12									0,7883371	0,6295301	0,6809288	0,3651402	0,2714356	0,2782449
13									0,7211194	0,5545722	0,5525494	0,2708744	0,27549	0,2733771

NOTA 1 – El valor SRC = 6, HRC = 5 (\*) se suprimió del análisis por superar los requisitos de registro temporal del plan de pruebas.

CUADRO 24

Datos subjetivos para todas las combinaciones HRC-SRC 625/50 – (valores de DMOS)

SRC (Imagen)	HRC									
	1	2	3	4	5	6	7	8	9	10
1		0,59461	0,64436	0,40804		0,34109		0,2677		0,26878
2		0,54173	0,70995	0,27443		0,22715		0,21133		0,16647
3		0,73314	0,76167	0,49848		0,38613		0,34574		0,26701
4		0,58528	0,90446	0,62361		0,61143		0,43329		0,26548
5		0,61973	0,68987	0,41648		0,4218		0,27543		0,2022
6		0,38852	0,44457	0,27983		0,28106		0,23726		0,17793
7				0,59953		0,55093			0,45163	0,35617
8				0,32528		0,32727			0,30303	0,26366
9				0,47656		0,49924			0,39101	0,37122
10				0,70492		0,58218			0,49711	0,37854
11	0,79919				0,59256		0,34337			0,30567
12	0,61418				0,6661		0,53242			0,44737
13	0,74225				0,66799		0,42065			0,33381



CUADRO 25

Datos subjetivos para todas las combinaciones HRC-SRC 525/60 – (valores de errores estándar)

SRC (Imagen)	HRC													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0,02109499	0,0223858	0,0202654	0,0200377										
2	0,02072424	0,0186353	0,0164296	0,0179823										
3	0,02075164	0,021336	0,0131301	0,0141977										
4					0,0224479	0,0200094	0,0221945	0,0216022						
5					0,0254351	0,0217278	0,0179396	0,0145813						
6						0,0215159	0,0176766	0,0180308						
7					0,0197204	0,0171224	0,0147712	0,0188843						
8									0,010892	0,0180687	0,0185947	0,0249537	0,0272349	0,0258362
9									0,0167711	0,018702	0,0281708	0,0226776	0,0193788	0,0203533
10									0,0144376	0,0263593	0,0171287	0,0202314	0,01996	0,018688
11									0,0186046	0,0189571	0,0213137	0,0188185	0,020292	0,0183653
12									0,0175106	0,0223805	0,0216039	0,0192717	0,0183	0,0202472
13									0,0213225	0,023069	0,0238845	0,0196748	0,0187747	0,0201108

NOTA 1 – Para convertirlos en desviaciones típicas, multiplicar por la raíz cuadrada del número de observaciones, 66.

NOTA 2 – El valor SRC = 6, HRC = 5 se suprimió del análisis por superar los requisitos de registro temporal del plan de pruebas.

## CUADRO 26

**Datos subjetivos para todas las combinaciones HRC-SRC 625/60 –  
(valores de los errores estándar)**

SRC (Imagen)	HRC									
	1	2	3	4	5	6	7	8	9	10
1		0,040255	0,039572	0,038567		0,040432		0,040014		0,036183
2		0,038683	0,033027	0,040957		0,038301		0,042618		0,033956
3		0,039502	0,039111	0,039109		0,042553		0,044151		0,036685
4		0,031762	0,024408	0,036375		0,031371		0,02973		0,042911
5		0,034299	0,044757	0,0407		0,03597		0,033742		0,041272
6		0,040602	0,040035	0,03707		0,043341		0,035289		0,040621
7				0,037894		0,032156		0,038034		0,036946
8				0,036819		0,041563		0,036988		0,037467
9				0,040289		0,040265		0,04015		0,039649
10				0,030283		0,038334		0,037966		0,041339
11	0,034761				0,034838		0,041778			0,041516
12	0,037332				0,036964		0,031253			0,035114
13	0,035205				0,038385		0,038371			0,043687

NOTA – Para convertirlos en desviaciones típicas, multiplicar por la raíz cuadrada del número de observaciones, 27.

## 6 Referencias Bibliográficas

ATIS [octubre de 2001] Technical Report T1.TR.72-2001 – Methodological Framework for Specifying Accuracy and Cross-Calibration of Video Quality Metrics, Alliance for Telecommunications Industry Solutions, 1200 G Street, NW Suite 500, Washington DC.