



UNIÓN INTERNACIONAL DE TELECOMUNICACIONES

**UIT-T**

SECTOR DE NORMALIZACIÓN  
DE LAS TELECOMUNICACIONES  
DE LA UIT

**J.144**

(03/2004)

SERIE J: REDES DE CABLE Y TRANSMISIÓN DE  
PROGRAMAS RADIOFÓNICOS Y TELEVISIVOS,  
Y DE OTRAS SEÑALES MULTIMEDIOS

Mediciones de la calidad de servicio

---

**Técnicas de medición objetiva de la percepción  
de la calidad vídeo en televisión digital por  
cable en presencia de una referencia completa**

Recomendación UIT-T J.144

---



## **Recomendación UIT-T J.144**

### **Técnicas de medición objetiva de la percepción de la calidad vídeo en televisión digital por cable en presencia de una referencia completa**

#### **Resumen**

Esta Recomendación contiene directrices sobre la selección del equipo apropiado de medición objetiva de la percepción de la calidad vídeo, diseñado para su utilización en aplicaciones de televisión digital por cable cuando se dispone de la señal de vídeo de referencia completa. El material de la prueba de validación no contenía errores de canal. En la presente Recomendación se definen modelos computacionales objetivos que han demostrado ser mejores que la (relación señal ruido de cresta, (PSNR, *peak signal to noise ratio*) como instrumentos automáticos de medida para evaluar la calidad de vídeo.

En esta revisión de la Recomendación se presentan cuatro métodos objetivos para la evaluación de la percepción de la calidad de vídeo en su sección normativa.

#### **Orígenes**

La Recomendación UIT-T J.144 fue aprobada el 15 de marzo de 2004 por la Comisión de Estudio 9 (2001-2004) del UIT-T por el procedimiento de la Recomendación UIT-T A.8.

## PREFACIO

La UIT (Unión Internacional de Telecomunicaciones) es el organismo especializado de las Naciones Unidas en el campo de las telecomunicaciones. El UIT-T (Sector de Normalización de las Telecomunicaciones de la UIT) es un órgano permanente de la UIT. Este órgano estudia los aspectos técnicos, de explotación y tarifarios y publica Recomendaciones sobre los mismos, con miras a la normalización de las telecomunicaciones en el plano mundial.

La Asamblea Mundial de Normalización de las Telecomunicaciones (AMNT), que se celebra cada cuatro años, establece los temas que han de estudiar las Comisiones de Estudio del UIT-T, que a su vez producen Recomendaciones sobre dichos temas.

La aprobación de Recomendaciones por los Miembros del UIT-T es el objeto del procedimiento establecido en la Resolución 1 de la AMNT.

En ciertos sectores de la tecnología de la información que corresponden a la esfera de competencia del UIT-T, se preparan las normas necesarias en colaboración con la ISO y la CEI.

## NOTA

En esta Recomendación, la expresión "Administración" se utiliza para designar, en forma abreviada, tanto una administración de telecomunicaciones como una empresa de explotación reconocida de telecomunicaciones.

La observancia de esta Recomendación es voluntaria. Ahora bien, la Recomendación puede contener ciertas disposiciones obligatorias (para asegurar, por ejemplo, la aplicabilidad o la interoperabilidad), por lo que la observancia se consigue con el cumplimiento exacto y puntual de todas las disposiciones obligatorias. La obligatoriedad de un elemento preceptivo o requisito se expresa mediante las frases "tener que, haber de, hay que + infinitivo" o el verbo principal en tiempo futuro simple de mandato, en modo afirmativo o negativo. El hecho de que se utilice esta formulación no entraña que la observancia se imponga a ninguna de las partes.

## PROPIEDAD INTELECTUAL

La UIT señala a la atención la posibilidad de que la utilización o aplicación de la presente Recomendación suponga el empleo de un derecho de propiedad intelectual reivindicado. La UIT no adopta ninguna posición en cuanto a la demostración, validez o aplicabilidad de los derechos de propiedad intelectual reivindicados, ya sea por los miembros de la UIT o por terceros ajenos al proceso de elaboración de Recomendaciones.

En la fecha de aprobación de la presente Recomendación, la UIT ha recibido notificación de propiedad intelectual, protegida por patente, que puede ser necesaria para aplicar esta Recomendación. Sin embargo, debe señalarse a los usuarios que puede que esta información no se encuentre totalmente actualizada al respecto, por lo que se les insta encarecidamente a consultar la base de datos sobre patentes de la TSB.

© UIT 2009

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

# ÍNDICE

	<b>Página</b>
1 Alcance .....	1
1.1 Aplicación.....	1
1.2 Limitaciones .....	1
2 Referencias .....	2
2.1 Referencias normativas .....	2
2.2 Referencias informativas .....	2
3 Términos, definiciones y siglas .....	2
4 Requisitos de usuario.....	3
5 Descripción del método de medición con referencia completa.....	3
6 Conclusiones del Grupo de expertos en calidad vídeo (VQEG) .....	4
7 Conclusiones.....	6
7.1 Recomendación general.....	6
7.2 Modelos objetivos de calidad de vídeo – El camino hacia futuras revisiones.....	6
Anexo A – British Telecommunications plc – Descripción funcional de un modelo de calidad de vídeo con referencia completa.....	7
A.1 Introducción.....	7
A.2 Modelo de referencia completa de BT .....	7
A.3 Detectores .....	7
A.4 Integración.....	16
A.5 Registro.....	17
A.6 Referencias .....	17
A.7 Datos subjetivos y objetivos.....	17
Anexo B – Yonsei University/SK Telecom/Radio Research Laboratory – Descripción funcional de un modelo de calidad de vídeo con referencia completa .....	22
B.1 Introducción.....	22
B.2 Medición objetiva de la calidad de vídeo a partir de la degradación de los bordes .....	22
B.3 Registro.....	33
B.4 Conclusión.....	36
B.5 Referencias .....	36
Anexo C – Telecommunications Research and Development Center (CPqD) – Descripción técnica de la evaluación de imágenes basada en segmentación (IES) .....	41
C.1 Introducción.....	41
C.2 Descripción general del algoritmo CPqD-IES .....	41
C.3 Corrección del desplazamiento y de la ganancia.....	43
C.4 Segmentación de la imagen.....	45

	<b>Página</b>
C.5	Medición objetiva..... 47
C.6	Base de datos de los modelos de degradación..... 47
C.7	Estimación de los modelos de degradación..... 50
C.8	Referencias ..... 52
C.9	Resultados objetivos de las pruebas VQEG-fase II..... 53
<b>Anexo D – Administración nacional de telecomunicaciones e información (NTIA, <i>national telecommunications and information administration</i>) – Descripción técnica de la métrica de la calidad de vídeo (VQM, <i>video quality metric</i>) .....</b>	
D.1	Introducción..... 54
D.2	Referencias ..... 54
D.3	Definiciones..... 55
D.4	Esquema del cálculo de la métrica de calidad del vídeo (VQM) ..... 59
D.5	Muestreo ..... 59
D.6	Calibración ..... 62
D.7	Características de calidad ..... 85
D.8	Parámetros de calidad..... 92
D.9	Modelo general..... 101
D.10	Referencias informativas ..... 102
D.11	Datos objetivos brutos de la métrica de la calidad de vídeo (VQM)..... 103
<b>Apéndice I – KDDI Esquema de evaluación objetiva de la calidad de vídeo y valoración de la calidad de funcionamiento .....</b>	
I.1	Alcance ..... 107
I.2	Esquema de evaluación de la calidad vídeo objetiva ..... 108
I.3	Implementación ..... 110
I.4	Resultados de la verificación ..... 112
<b>Apéndice II – Tektronix Inc. y Sarnoff Corporation – Medición objetiva de la calidad de percepción de vídeo utilizando una técnica de referencia completa basada en JND...</b>	
II.1	Alcance, objeto y aplicación..... 115
II.2	Referencias ..... 121
II.3	Introducción..... 121
II.4	Presentación del algoritmo ..... 123
II.5	Detalles del algoritmo..... 128
<b>Apéndice II – Adjunto 1 – Bibliografía .....</b>	
<b>Apéndice II – Adjunto 2 – Factores de prueba, tecnologías de codificación y aplicaciones...</b>	
<b>Apéndice II – Adjunto 3 – Clasificación de errores.....</b>	

## Introducción

La televisión digital da lugar a nuevas consideraciones con respecto a la calidad de servicio, con complejas relaciones entre las mediciones objetivas de los parámetros y la calidad subjetiva de la imagen. Si bien es conveniente efectuar las mediciones objetivas con una buena correlación con la evaluación subjetiva de la calidad para obtener una calidad de servicio óptima en el funcionamiento de los sistemas de televisión por cable, cabe recordar que las mediciones objetivas no sustituyen directamente la evaluación subjetiva de la calidad.

Las evaluaciones subjetivas de la calidad son procedimientos cuidadosamente diseñados para determinar la opinión promedio de observadores humanos sobre una serie de secuencias de vídeo para una aplicación determinada. Los resultados de esas pruebas se utilizan en la preparación de sistemas básicos y evaluaciones de referencia. Las evaluaciones subjetivas de la calidad para una aplicación diferente, en condiciones de prueba distintas, seguirán proporcionando resultados significativos, aunque probablemente difieran las notas de opinión sobre la misma serie de secuencias de vídeo. Las mediciones objetivas se utilizan en una amplia gama de aplicaciones produciendo los mismos resultados con un conjunto determinado de secuencias de vídeo. La selección de las secuencias de vídeo que se han de utilizar y la interpretación de las mediciones objetivas resultantes son algunos de los factores que varían para una aplicación específica.

Las mediciones objetivas y la evaluación subjetiva de la calidad son, por tanto, complementarias pero no intercambiables. Si la evaluación subjetiva se utiliza con fines relacionados con la investigación, las mediciones objetivas se necesitan para las especificaciones de equipos y la medición y el control diarios de la calidad de funcionamiento de los sistemas.

Para los fines de esta Recomendación se han adoptado las siguientes definiciones.

- "Evaluación subjetiva", que se refiere a la determinación de la calidad o el deterioro de imágenes de tipo programa presentadas a un grupo de evaluadores humanos en sesiones de observación visual.
- "Medición perceptiva objetiva", que se refiere a la medición de las características de una cadena de programa utilizando imágenes de tipo programa y métodos de medición objetiva (con instrumentos) para obtener un indicio que se acerque a la nota que se obtendría con una prueba de evaluación subjetiva.
- "Medición de la señal", que se refiere a la medición de la calidad de una cadena de programa utilizando señales de prueba y métodos de medición objetiva (con instrumentos).

En la presente Recomendación los términos medición objetiva y medición perceptiva se utilizan indistintamente para referirse a la medición perceptiva objetiva.

Hay tres métodos básicos para efectuar las mediciones objetivas:

- FR – Método aplicable cuando se dispone de la señal de vídeo de referencia completa. Se trata del método de dos extremos, del cual se ocupa la presente Recomendación.
- RR – Método aplicable cuando sólo se dispone de información de referencia de vídeo reducida. Se trata también de un método de dos extremos que es objeto de una Recomendación aparte [en estudio].
- NR – Método aplicable cuando no se dispone de señal de vídeo de referencia ni de información. Se trata de un método de un solo extremo y es objeto de una Recomendación aparte [en estudio].

Cada método tiene sus propias aplicaciones, y proporciona grados diferentes de exactitud de la medición, expresada en términos de correlación con los resultados de la evaluación subjetiva.



## Recomendación UIT-T J.144

### Técnicas de medición objetiva de la percepción de la calidad vídeo en televisión digital por cable en presencia de una referencia completa

#### 1 Alcance

Esta Recomendación contiene directrices sobre la selección del equipo apropiado de medición de la percepción de la calidad vídeo, diseñado para su utilización en aplicaciones de televisión digital por cable cuando se puede utilizar el método de medición con referencia completa.

El método de medición de referencia completa puede utilizarse cuando la señal de vídeo de referencia no degradada está fácilmente disponible en el punto de medición, como puede ser el caso de las mediciones en equipos individuales o en una cadena en el laboratorio o en un entorno cerrado, por ejemplo, el extremo de cabecera de la televisión por cable. Los métodos de estimación se basan en el procesamiento de vídeo de componente digital de 8 bits con arreglo a la definición de la Rec. UIT-R BT.601-5<sup>1</sup>. El codificador puede utilizar varios métodos de compresión (por ejemplo, MPEG, H.263, etc.). Los modelos propuestos en la presente Recomendación pueden utilizarse para evaluar un códec (combinación de codificador y decodificador) o una concatenación de varios métodos de compresión y dispositivos de almacenamiento de memoria. Aunque la obtención de los estimadores de calidad objetivos descritos en la presente Recomendación puede haber tenido en cuenta la degradación por error (por ejemplo, errores en los bits, paquetes rechazados, etc.), no hay resultados de pruebas independientes que permitan validar actualmente, la utilización de estimadores en sistemas con degradación por error. El material de la prueba de validación no contenía errores de canal. Contenía sin embargo, degradaciones de codificación y velocidades de compresión entre 768 kbit/s y 5 Mbit/s.

#### 1.1 Aplicación

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

- 1) la evaluación, especificación y prueba de aceptación de códecs, congruente con la precisión limitada descrita más adelante;
- 2) la supervisión potencial de la calidad en tiempo real y en servicio, en la fuente;
- 3) la supervisión a distancia de la calidad de destino cuando hay disponible una copia de la fuente;
- 4) la medición de la calidad de un sistema de transmisión o almacenamiento que utilice técnicas de compresión y descompresión de vídeo, ya sea de forma aislada o como combinación de dichas técnicas.

#### 1.2 Limitaciones

Los modelos de estimación descritos en esta Recomendación no pueden utilizarse para sustituir las pruebas subjetivas. Los valores de correlación entre dos pruebas subjetivas cuidadosamente

---

<sup>1</sup> Esto no impide la implementación del método de medida para sistemas de vídeo unidireccionales con entradas y salidas de vídeo compuesto. La especificación de la conversión entre los dominios compuesto y de componentes no forma parte de la presente Recomendación. SMTPE 170M, por ejemplo, especifica un método de ejecutar esta conversión para NTSC.

diseñadas y ejecutadas (es decir en dos laboratorios distintos) suelen quedar en el intervalo 0,92 a 0,97. En la presente Recomendación no se definen medios de cuantificar los posibles errores de estimación. Los usuarios de la presente Recomendación deben consultar la comparación de resultados disponibles, subjetivos y objetivos, para formarse una idea de la magnitud de los errores de estimación del índice de calidad del vídeo.

La calidad de funcionamiento prevista por los modelos de estimación no se ha validado aún para sistemas de vídeo con degradación por errores en el canal de transmisión.

## 2 Referencias

Las siguientes Recomendaciones del UIT-T y otras referencias contienen disposiciones que, mediante su referencia en este texto, constituyen disposiciones de la presente Recomendación. Al efectuar esta publicación, estaban en vigor las ediciones indicadas. Todas las Recomendaciones y otras referencias son objeto de revisiones por lo que se preconiza que los usuarios de esta Recomendación investiguen la posibilidad de aplicar las ediciones más recientes de las Recomendaciones y otras referencias citadas a continuación. Se publica periódicamente una lista de las Recomendaciones UIT-T actualmente vigentes. En esta Recomendación, la referencia a un documento, en tanto que autónomo, no le otorga el rango de una Recomendación.

### 2.1 Referencias normativas

- Recomendación UIT-R BT.601-5 (1995), *Parámetros de codificación de televisión digital para estudios con formatos de imagen normal 4:3 y de pantalla ancha 16:9.*

### 2.2 Referencias informativas

- Recomendación UIT-T J.140 (1998), *Evaluación subjetiva de la calidad de imagen para sistemas de televisión digital por cable.*
- Recomendación UIT-T J.143 (2000), *Requisitos de usuario para mediciones objetivas de la percepción de la calidad vídeo en televisión digital por cable.*
- Recomendación UIT-T P.910 (1996), *Métodos de evaluación subjetiva de la calidad vídeo para aplicaciones multimedios.*
- Material de referencia UIT-T (2004), *Objective perceptual assessment of video quality: Full reference television.*
- Recomendación UIT-T P.911 (1998), *Métodos de evaluación subjetiva de la calidad audiovisual para aplicaciones multimedios.*
- U. S. Standards Committee T1\* Technical Report T1.TR.73-2001, *Video normalization methods applicable to objective video quality metrics utilizing a full reference technique.*
- Recomendación UIT-R BT.500-11 (2002), *Metodología para la evaluación subjetiva de la calidad de las imágenes de televisión.*

## 3 Términos, definiciones y siglas

En esta Recomendación se definen los términos siguientes.

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

---

\* Las normas T1 son mantenidas por ATIS desde noviembre 2003.

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

**3.3 medición de la señal (imagen):** Medición de la calidad de una cadena de programa utilizando señales de prueba y métodos de medición objetiva (con instrumentos).

**3.4 ANOVA:** Análisis de la varianza (*analysis of variance*).

**3.5 FRTV:** Televisión de referencia completa (*full reference television*).

**3.6 DSCQS:** Escala de calidad continua de doble estímulo (*double stimulus continuous quality scale*).

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

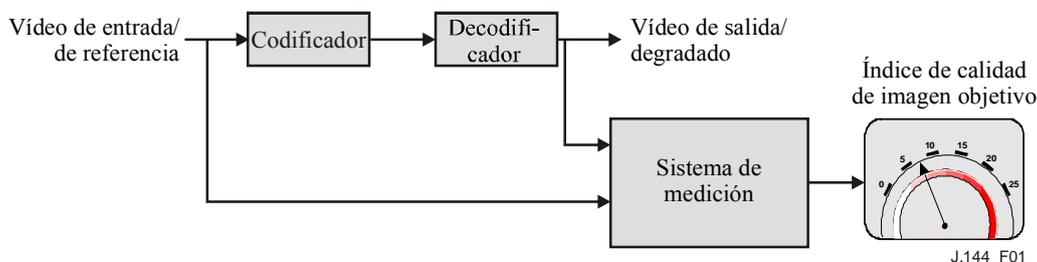
#### 4 Requisitos de usuario

Los requisitos de usuario para los métodos de medición perceptiva de la calidad de imagen se definen en la Rec. UIT-T J.143.

#### 5 Descripción del método de medición con referencia completa

El método de medición de dos extremos con referencia completa, para la medición objetiva de la calidad de vídeo percibida, evalúa la calidad de funcionamiento de los sistemas estableciendo una comparación entre la entrada no distorsionada, o referencia, que es la señal de vídeo a la entrada del sistema, y la señal degradada a la salida del sistema (Figura 1).

La Figura 1 muestra un ejemplo de aplicación del método con referencia total para probar un códec en laboratorio.



**Figura 1 – Aplicación del método de medición de la calidad percibida con referencia completa para la prueba de un códec en laboratorio**

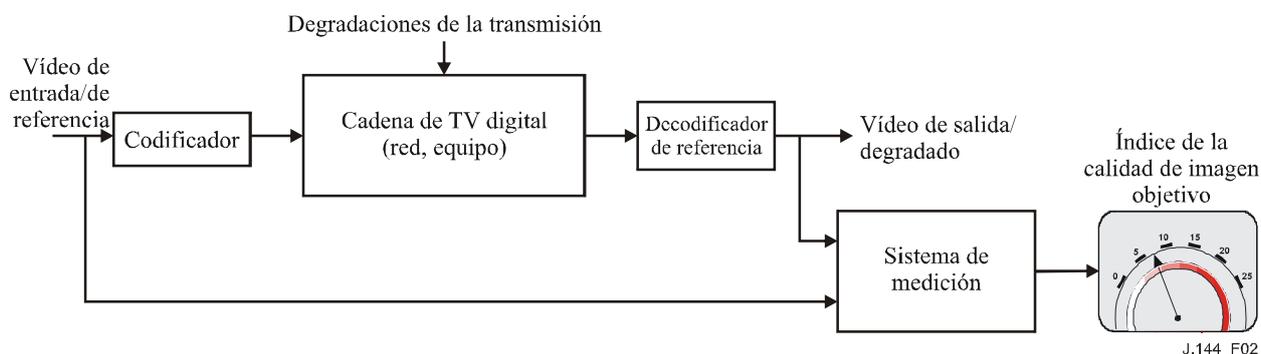
La comparación entre las señales de entrada y salida requiere un proceso de alineación espacial y temporal para compensar los posibles desplazamientos o recortes de la imagen vertical u horizontalmente. Quizás también se requiera corregir cualquier desplazamiento o diferencia de ganancia en los canales de luminancia y crominancia. A continuación se calcula el índice objetivo de la calidad de la imagen, aplicando normalmente un modelo perceptivo de la visión humana.

Se denomina normalización al ajuste de la alineación y de la ganancia. Este proceso es necesario porque la mayor parte de los métodos de referencia completa comparan efectivamente la imagen de referencia y la procesada píxel a píxel. Un ejemplo puede ser el cálculo de la relación señal/ruido de cresta (PSNR, *peak signal to noise ratio*). Sólo se suprimen del vídeo las variaciones estáticas invariables en el tiempo, las variaciones dinámicas generadas por el sistema que se prueba se miden como parte del cálculo del índice de calidad. Se puede encontrar una exposición detallada de los

motivos de normalización en el U.S. Standards Committee T1\* Technical Report T1.TR.73-2001, "Video Normalization Methods Applicable to Objective Video Quality Metrics Utilizing a Full Reference Technique". Las métricas de la calidad de vídeo definidas en los anexos A a D incluyen métodos de normalización asociados. Pueden utilizarse otros métodos de normalización para las métricas de calidad de vídeo que figuran en los anexos A a D y en los apéndices I y II, con tal que proporcionen la precisión de normalización necesaria.

Dado que las métricas de la calidad de vídeo se suelen basar en aproximaciones a la respuesta visual humana, y no en la medición de artefactos de codificación específicos, resultan en principio igualmente válidas para los sistemas analógicos y los digitales. También son válidas para las cadenas en las que se combinan sistemas analógicos y digitales, o en las que hay sistemas de compresión digital concatenados.

La Figura 2 muestra un ejemplo de aplicación del método de referencia completa para la prueba de una cadena de transmisión.



**Figura 2 – Aplicación del método de medición de la calidad percibida con referencia completa para la prueba de una cadena de transmisión**

En este caso, la entrada al decodificador de referencia procede de diversos puntos de la cadena de transmisión; el decodificador puede estar situado, por ejemplo, en un punto de la red, como en la Figura 2, o directamente a la salida del codificador como en la Figura 1. Si la cadena de transmisión digital es transparente, la medición del índice de calidad de imagen objetivo en la fuente es igual a la medición en cualquier punto subsiguiente de la cadena.

Por lo general, se acepta que con el método de referencia completa se obtiene la mayor exactitud de las mediciones de la calidad de imagen percibida. El método ha demostrado la posibilidad de establecer un alto grado de correlación con las evaluaciones subjetivas efectuadas de conformidad con la especificación de la escala de calidad continua de doble estímulo (DSCQS) de UIT-R BT.500-11.

## 6 Conclusiones del Grupo de expertos en calidad vídeo (VQEG)

Los estudios sobre la medición de la calidad vídeo percibida se llevan a cabo en un grupo oficioso, llamado Grupo de expertos en calidad vídeo (VQEG, *video quality experts group*), que informa de sus resultados a la Comisión de Estudio 9 del UIT-T y a la Comisión de Estudio 6 del UIT-R. La primera tarea del VQEG consistió en evaluar el rendimiento de los algoritmos propuestos para la medición de la calidad percibida de vídeo con referencia completa.

El VQEG publicó un extenso proyecto de informe final sobre la primera fase de su trabajo en marzo de 2000. El VQEG publicó el informe final de la fase II de la prueba de televisión con referencia completa en agosto de 2003.

Se invita al lector a estudiar dicho Informe para adquirir un conocimiento profundo del trabajo efectuado por el VQEG. La finalidad de éste era comprobar los modelos propuestos en términos de:

- exactitud de la predicción (capacidad del modelo de predecir la calidad subjetiva);
- monotonía de la predicción (grado de concordancia de las predicciones del modelo con la clasificación por categorías de los índices de calidad subjetivos);
- coherencia de la predicción (grado en que el modelo mantiene la exactitud de la predicción en la gama de secuencias de prueba de vídeo y sistemas de vídeo, es decir, en qué medida su respuesta se mantiene con diversas degradaciones de vídeo).

La fase I de la prueba de validación de la FRTV del VQEG no proporcionó pruebas suficientes que permitieran identificar un método recomendable para la evaluación objetiva de la calidad de la percepción de vídeo. Los resultados de la fase II de la prueba de validación de la FRTV del VQEG indicaron que había cuatro métodos adecuados para su inclusión en la parte normativa de la presente Recomendación.

En la fase II, la evaluación subjetiva se abstuvo estrictamente al método de escala de calidad continua de doble estímulo (DSCQS, *double stimulus continuous quality scale*) de la BT.500-11. En los planes de pruebas subjetivas y objetivas hay procedimientos para el análisis de la validación de las notas subjetivas, y cuatro métricas para la comparación de los datos objetivos con los resultados subjetivos. Se hicieron además otros análisis estadísticos, entre ellos el ANOVA (análisis de la varianza) y la prueba F.

De acuerdo con estas pruebas, se pueden recomendar a la UIT cuatro métodos en este momento, a saber:

Anexo A – British Telecom (Reino Unido, proponente D del VQEG).

Anexo B – Yonsei University/SK Telecom/Radio Research Laboratory (República de Corea, proponente E del VQEG).

Anexo C – CPqD (República Federal de Brasil, proponente F del VQEG).

Anexo D – NTIA (Estados Unidos de América, proponente H del VQEG).

Las descripciones técnicas de estos modelos aparecen en los anexos A a D respectivamente. Obsérvese que el orden de los anexos es totalmente arbitrario y no facilita indicación alguna de la bondad de la predicción de la calidad. Por ejemplo, en términos absolutos el modelo de NTIA presenta la mejor correlación con los índices subjetivos de la prueba 525.

Los Cuadros 1 y 2 ofrecen detalles informativos sobre la bondad de los modelos en la Fase II de la prueba FRTV de la VQEG. Para los datos 525, los modelos NTIA y BT se comportaron estadísticamente mejor que los demás modelos, y son estadísticamente equivalentes entre sí. Para los datos 625, hay tres modelos (CPqD, NTIA, Yonsei/SKT/RRL) estadísticamente equivalentes entre sí y que a su vez son estadísticamente mejores que el otro modelo. Hay que subrayar asimismo que sólo el modelo NTIA consiguió empatar estadísticamente con los mejores comportamientos en ambas pruebas.

**Cuadro 1 – Descripción informativa del comportamiento de los modelos en la fase II de la prueba FRTV del VQEG (datos 525)**

Métrica	BT	Yonsei/SKT/RRL	CPqD	NTIA	PSNR (Nota)
Anexo	A	B	C	D	
Correlación de Pearson	0,937	0,857	0,835	0,938	0,804
Error RMS	0,075	0,110	0,117	0,074	0,127
NOTA – Los valores de PSNR aquí indicados se han obtenido del informe final de la fase II del VQEG. Estos valores se calcularon en Yonsei.					

**Cuadro 2 – Descripción informativa del comportamiento de los modelos en la fase II de la prueba FRTV del VQEG (datos 625)**

<b>Métrica</b>	<b>BT</b>	<b>Yonsei/SKT/RRL</b>	<b>CPqD</b>	<b>NTIA</b>	<b>PSNR</b>
Anexo	A	B	C	D	
Correlación de Pearson	0,779	0,870	0,898	0,886	0,733
Error RMS	0,113	0,089	0,079	0,083	0,122

## **7 Conclusiones**

### **7.1 Recomendación general**

Cuando se efectúen mediciones de la calidad vídeo percibida, utilizando el modelo de referencia completa descrito en esta Recomendación, los operadores deberán analizar primero la equivalencia de sus requisitos específicos de aplicación y usuario en términos de características y calidad de funcionamiento del equipo de medición.

A continuación se indican algunos aspectos que se han de tomar en consideración:

- costo imputable a la propiedad del equipo de medición perceptiva;
- servicio posventa del fabricante;
- facilidad de funcionamiento;
- fiabilidad;
- tamaño, peso y requisitos de alimentación;
- rapidez de la medición en tiempo real y en diferido;
- funcionamiento en línea (en servicio);
- exactitud, monotonía y coherencia de la predicción.

Se recomiendan los cuatro métodos normativos por su elevada correlación con los resultados objetivos de la fase II de las pruebas FRTV del VQEG. No obstante, hasta que estos métodos estén disponibles en equipos comerciales de prueba, se recomiendan los métodos Tektronix/Sarnoff y KDDI. Estas dos métricas tradicionales de la calidad de vídeo se especifican en los apéndices I y II. Se han validado en la fase I del VQEG y se incluyen debido a su alto grado de utilización en el importante parque de instrumentos de medición de la calidad de vídeo instalado. La métrica de calidad de vídeo probada en la fase I del VQEG no se consideró suficientemente precisa para incluirse como norma en una Recomendación de la UIT. No obstante, debe tenerse en cuenta que en la fase I del VQEG las condiciones de prueba fueron mucho más diversas, muchas de ellas con degradaciones muy pequeñas comparadas con a las empleadas en la fase II del VQEG. En la fase II del VQEG se utilizaron condiciones de prueba que empleaban la transformada discreta de coseno (es decir MPEG-2 y H.263) con velocidades binarias entre 768 kbit/s y 5 Mbit/s.

### **7.2 Modelos objetivos de calidad de vídeo – El camino hacia futuras revisiones**

Para que un modelo se convierta en norma, debe ser comprobado por un órgano abierto e independiente (como el VQEG) que efectúe la evaluación técnica en base a las directrices y criterios de calidad establecidos por la Comisión de Estudio 9 (CE 9). Lo que pretende a la postre la CE 9 es recomendar un solo método normativo de referencia completa para la televisión por cable.

## Anexo A

### British Telecommunications plc

#### Descripción funcional de un modelo de calidad de vídeo con referencia completa

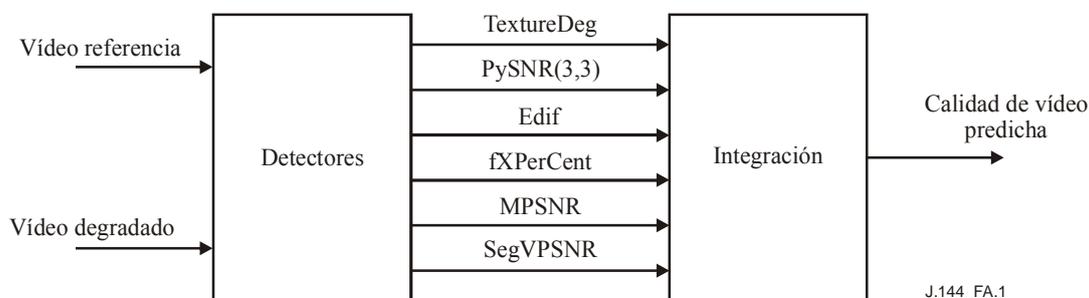
##### A.1 Introducción

El instrumento de evaluación automática de la calidad de vídeo con 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 un programa informático que implementaba el modelo, y la calidad de funcionamiento obtenida se presentó en el informe de la prueba [A-1].

##### A.2 Modelo de referencia completa de BT

El algoritmo BTFR consiste en una detección seguida de una integración como se muestra en la Figura A.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.

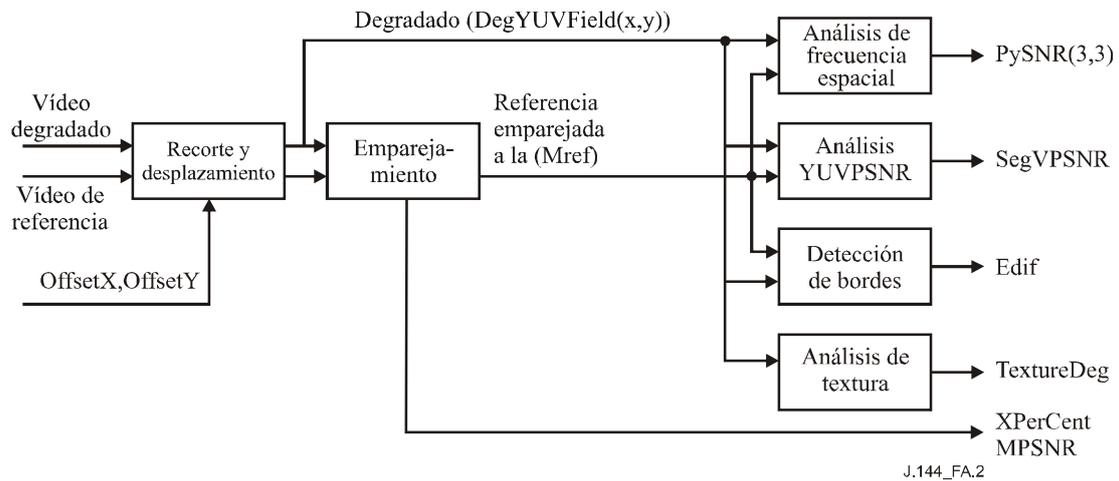


**Figura A.1 – Modelo de evaluación de la calidad de vídeo de referencia completa**

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

##### A.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 Figura A.2.



**Figura A.2 – Detección**

### A.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  $RefV$ :

$$Re fY(x, y) \quad x = 0..X - 1, \quad y = 0..Y - 1 \quad (A.3.1-1)$$

$$Re fU(x, y) \quad x = 0..X - 1, \quad y = 0..Y - 1 \quad (A.3.1-2)$$

$$Re fV(x, y) \quad x = 0..X - 1, \quad y = 0..Y - 1 \quad (A.3.1-3)$$

Siendo  $X$  el número de píxeles horizontales de un campo e  $Y$  el número de píxeles verticales. Para una entrada  $YUV422$ , debe repetirse cada valor  $U$  y  $V$  para obtener las ecuaciones A.3.1-2 y A.3.1-3 de resolución completa.

### A.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  $OffsetX$  y  $OffsetY$  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 izquierda 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 A.3.1) en las matrices  $InYField$ ,  $InUField$ , e  $InVField$ , la salida recortada y desplazada se calcula mediante las ecuaciones A.3.2-1 a A.3.2-17.

$$XStart = -OffsetX \quad (A.3.2-1)$$

$$si (XStart < C_x) \text{ entonces } XStart = C_x \quad (A.3.2-2)$$

$$XEnd = X - 1 - OffsetX \quad (A.3.2-3)$$

$$si (XEnd > X - C_x - 1) \text{ entonces } XEnd = X - C_x - 1 \quad (A.3.2-4)$$

$$YStart = -OffsetY \quad (A.3.2-5)$$

$$si (YStart < C_y) \text{ entonces } YStart = C_y \quad (A.3.2-6)$$

$$YEnd = Y - 1 - OffsetY \quad (A.3.2-7)$$

$$si (YEnd > Y - C_y - 1) entonces \quad YEnd = Y - C_y - 1 \quad (A.3.2-8)$$

$X$  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,

$$X = 720, \quad Y = 288, \quad C_x = 30, \quad C_y = 10 \quad (A.3.2-9)$$

Para las secuencias de 525,

$$X = 720, \quad Y = 243, \quad C_x = 30, \quad C_y = 10 \quad (A.3.2-10)$$

$Xstart$ ,  $Xend$ ,  $Ystart$  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 A.3.2-11 a A.3.2-12, siendo  $YField$ ,  $UField$  y  $VField$  las 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:

$$YField(x, y) = 0 \quad x = 0..XStart - 1, XEnd + 1..X - 1 \quad y = 0..Y - 1 \quad (A.3.2-11)$$

$$UField(x, y) = VField(x, y) = 128 \quad x = 0..XStart - 1, XEnd + 1..X - 1 \quad y = 0..Y - 1 \quad (A.3.2-12)$$

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

$$YField(x, y) = 0 \quad x = XStart..XEnd, \quad y = 0..YStart - 1, YEnd + 1..Y - 1 \quad (A.3.2-13)$$

$$UField(x, y) = VField(x, y) = 128 \quad x = XStart..XEnd \quad y = 0..YStart - 1, YEnd + 1..Y - 1 \quad (A.3.2-14)$$

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

$$YField(x, y) = InYField(x + OffsetX, y + OffsetY) \quad x = XStart..XEnd \quad y = YStart..YEnd \quad (A.3.2-15)$$

$$UField(x, y) = InUField(x + OffsetX, y + OffsetY) \quad x = XStart..XEnd \quad y = YStart..YEnd \quad (A.3.2-16)$$

$$VField(x, y) = InVField(x + XOffset, y + YOffset) \quad x = XStart..XEnd \quad y = YStart..YEnd \quad (A.3.2-17)$$

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.

### A.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:

$$Ref(N, Px, Py) = RefYField(Px, Py) \quad del \ campo \ N \quad (A.3.3-1)$$

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

$$Block\ Re\ f(N, Px, Py) = Re\ f(n, x, y) \quad x = Px - 4..Px + 4, \quad y = Py - 4..Py + 4 \quad (A.3.3-2)$$

$Deg(n, x, y)$  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:

$$E(N, Px, Py) = MIN((1/81) \sum_{j=-4}^4 \sum_{k=-4}^4 (Deg(N, Px + j, Py + k) - Re\ f(n, x + j, y + k))^2) \quad (A.3.3-3)$$

con:

$$\begin{aligned} n &= N - 4, \dots, N + 5 \\ x &= Px - 4, Px, \dots, Px + 4 \\ y &= Py - 4, Py, \dots, Py + 4 \end{aligned}$$

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

Si mediante la ecuación A.3.3-3 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:

$$M\ Re\ f(N, Px + j, Py + k) = Re\ f(n_m, x_m + j, y_m + k) \quad j = -4..4, k = -4..4 \quad (A.3.3-4)$$

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:

$$Px(h) = 16 + 8 \times h \quad h = 0..Qx - 1 \quad (A.3.3-5)$$

y

$$Py(v) = 16 + 8 \times v \quad v = 0..Qy - 1 \quad (A.3.3-6)$$

donde  $Qx$  y  $Qy$  definen el número de bloques de análisis horizontales y verticales. Debido a que el proceso de emparejamiento se ejecuta en base a bloques de  $9 \times 9$  píxeles, los bloques de emparejamiento adyacentes se solapan en un píxel.  $Mref$  se actualiza de acuerdo con las ecuaciones A.3.3-4, A.3.3-5 y A.3.3-6 por lo que las regiones de solapamiento de  $Mref$  se sobrescriben con los resultados de los cálculos subsiguientes.

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

$$BlockM\ Re\ f(N, Px(h), Py(v)) \quad h = 0..Qx - 1, \quad v = 0..Qy - 1 \quad (A.3.3-7)$$

y un conjunto de valores de error del mejor emparejamiento:

$$E(N, Px(h), Py(v)) \quad h = 0..Qx - 1, \quad v = 0..Qy - 1 \quad (A.3.3-8)$$

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

$$\begin{aligned} BlockM\ Re\ f(N, Px(h), Py(v)) &= Block\ Re\ f(MatT(h, v), MatX(h, v), MatY(h, v)) \\ & \quad h = 0..Qx - 1, \quad v = 0..Qy - 1 \end{aligned} \quad (A.3.3-9)$$

Los parámetros de emparejamiento para secuencias de radiodifusión de 625 y 525 se presentan en el Cuadro A.1.

**Cuadro A.1 – Parámetros de búsqueda  
para el procedimiento de emparejamiento**

Parámetro	625	525
$Q_x$	87	87
$Q_y$	33	28

La región de análisis definida por las ecuaciones A.3.3-6 y A.3.3-7 no ocupa todo el tamaño del campo. Por ello debe inicializarse  $MRef$  con arreglo a la ecuación A.3.3-9 para poder utilizarla sin restricciones en cualquier sitio.

$$MRef(x, y) = 0 \quad x = 0..X-1, \quad y = 0..Y-1 \quad (A.3.3-10)$$

### A.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 A.3.3-3, se utiliza en la construcción del histograma  $histX$  para cada campo con arreglo a la siguiente expresión:

$$histX(MatX(h, v) - Px(h) + 4) = histX(MatX(h, v) - Px(h) + 4) + 1 \quad (A.3.3.1-1)$$

$$h = 0..Q_x - 1, \quad v = 0..Q_y - 1$$

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 expresión:

$$fXPerCent = 100 \times \frac{Max(histX(i))}{\sum_{j=0}^8 histX(j)} \quad i = 0..8 \quad (A.3.3.2-2)$$

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

### A.3.3.2 PSNR emparejada

Se utiliza el error mínimo,  $E()$ , de cada bloque emparejado para calcular la relación señal-ruido emparejada con arreglo a las siguientes expresiones:

$$si \left( \sum_{h=0}^{Q_x-1} \sum_{v=0}^{Q_y-1} E(N, Px(h), Py(v)) > 0 \right) \text{ entonces} \quad (A.3.3.2-1)$$

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

$$si \left( \sum_{h=0}^{Q_x-1} \sum_{v=0}^{Q_y-1} E(N, Px(h), Py(v)) = 0 \right) \text{ entonces } MPSNR = 10 \log_{10} (255^2) \quad (A.3.3.2-2)$$

### A.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:

$$SyncT(h, v) = MatT(h, v) - N \quad h = 0..Q_x - 1, \quad v = 0..Q_y - 1 \quad (A.3.3.3-1)$$

$$SyncX(h, v) = MatX(h, v) - Px(h) \quad h = 0..Q_x - 1, \quad v = 0..Q_y - 1 \quad (A.3.3.3-2)$$

$$SyncY(h, v) = MatY(h, v) - Py(h) \quad h = 0..Q_x - 1, \quad v = 0..Q_y - 1 \quad (A.3.3.3-3)$$

### A.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 (véase la Figura A.3). 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 relación señal-ruido piramidal.

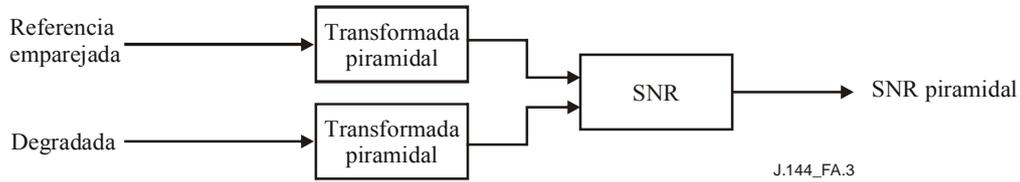


Figura A.3 – Análisis espacial de frecuencia

#### A.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:

$$P(x, y) = F(x, y) \quad x = 0..X-1, \quad y = 0..Y-1 \quad (\text{A.3.4.1-1})$$

A continuación se actualiza esta matriz piramidal en tres etapas (stage = 0..2) de análisis vertical y horizontal. El análisis horizontal  $H_{py}(\text{stage})$  viene definido por las ecuaciones A.3.4.1-2 a A.3.4.1-6.

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

$$PTemp(x, y) = P(x, y) \quad x = 0..X-1, \quad y = 0..Y-1 \quad (\text{A.3.4.1-2})$$

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

$$Tx = X / 2^{(\text{stage}+1)} \quad (\text{A.3.4.1-3})$$

$$Ty = Y / 2^{\text{stage}} \quad (\text{A.3.4.1-4})$$

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 la siguientes expresiones:

$$P(x, y) = 0.5 \times (PTemp(2x, y) + PTemp(2x+1, y)) \quad x = 0..Tx-1 \quad y = 0..Ty-1 \quad (\text{A.3.4.1-5})$$

$$P(x + Tx, y) = PTemp(2x, y) - PTemp(2x+1, y) \quad x = 0..Tx-1 \quad y = 0..Ty-1 \quad (\text{A.3.4.1-6})$$

El análisis vertical  $V_{py}(\text{stage})$  viene definido por las ecuaciones A.3.4.1-7 a A.3.4.1-11.

$$PTemp(x, y) = P(x, y) \quad x = 0..X-1, \quad y = 0..Y-1 \quad (\text{A.3.4.1-7})$$

$$Tx = X / 2^{\text{stage}} \quad (\text{A.3.4.1-8})$$

$$Ty = Y / 2^{(\text{stage}+1)} \quad (\text{A.3.4.1-9})$$

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 la siguientes expresiones:

$$P(x, y) = 0.5 \times (PTemp(x, 2y) + PTemp(x, 2y+1)) \quad x = 0..Tx-1, \quad y = 0..Ty-1 \quad (\text{A.3.4.1-10})$$

$$P(x, y + Ty) = PTemp(x, 2y) - PTemp(x, 2y+1) \quad x = 0..Tx-1 \quad y = 0..Ty-1 \quad (\text{A.3.4.1-11})$$

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..3) construidos con arreglo a la Figura A.4:

Q(0,0)	Q(0,1)	Q(0,0) = promedio de los bloques de 4
Q(0,2)	Q(0,3)	Q(0,1) = diferencia horizontal de los bloques de 4
		Q(0,2) = diferencia vertical de los bloques de 4
		Q(0,3) = diferencia diagonal de los bloques de 4

**Figura A.4 – Cuadrante de salida del análisis de la etapa 0**

A continuación se efectúa el análisis de la etapa 1 sobre Q(0,0) y los resultados Q(1,0..3) generados se almacenan en la pirámide con arreglo a la Figura A.5:

Q(1,0)	Q(1,1)	Q(0,1)
Q(1,2)	Q(1,3)	
Q(0,2)		Q(0,3)

**Figura A.5 – Cuadrante de salida del análisis de la etapa 1**

En el análisis de la etapa 2 se procesa Q(1,0) y se sustituye por Q(2,0..3).

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

#### A.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 expresión:

$$E(s, q) = (1/(XY)^2) \sum_{x=x1(s,q)}^{x2(s,q)-1} \sum_{y=y1(s,q)}^{y2(s,q)-1} (Pref(x, y) - Pdeg(x, y))^2 \quad s = 0..2 \quad q = 1..3 \quad (\text{A.3.4.2-1})$$

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

$$x1(s,1) = X / 2^{(s+1)} \quad x2(s,1) = 2 \times x1(s,1) \quad y1(s,1) = 0 \quad y2(s,1) = Y / 2^{(s+1)} \quad (\text{A.3.4.2-2})$$

$$x1(s,2) = 0 \quad x2(s,2) = X / 2^{(s+1)} \quad y1(s,2) = Y / 2^{(s+1)} \quad y2(s,2) = 2 \times y1(s,2) \quad (\text{A.3.4.2-3})$$

$$x1(s,3) = X / 2^{(s+1)} \quad x2(s,3) = 2 \times x1(s,3) \quad y1(s,3) = Y / 2^{(s+1)} \quad y2(s,3) = 2 \times y1(s,3) \quad (\text{A.3.4.2-4})$$

A continuación se utilizan los resultados de (A.3.4.2-1) para medir la PSNR de cada uno de los cuadrantes de cada campo con arreglo a la siguiente expresión:

$$si \ (E > 0.0) \quad PySNR(s, q) = 10.0 \times \log_{10}(255^2 / E(s, q)) \quad (\text{A.3.4.2-5})$$

$$entonces \ SNR = 10.0 \times \log_{10}(255^2 \times (XY)^2)$$

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

### A.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 A.3.5-1 a A.3.5-6.

Para cada campo se inicializa primero un contador de puntos de inflexión con arreglo a la ecuación A.3.5-1.

$$sum = 0 \quad (A.3.5-1)$$

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

$$last\_pos = 0, \quad last\_neg = 0 \quad (A.3.5-2)$$

$$dif(x) = P(x, y) - P(x + 1, y) \quad (A.3.5-3)$$

$$si((dif(x) < 0) AND (last\_neg < last\_pos)) then sum = sum + 1 \quad (A.3.5-4)$$

$$si((dif(x) > 0) AND (last\_neg > last\_pos)) then sum = sum + 1 \quad (A.3.5-5)$$

$$si(dif(x) > 0) then last\_pos = x \quad (A.3.5-6)$$

$$si(dif(x) < 0) then last\_neg = x \quad (A.3.5-7)$$

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:

$$TextureDeg = sum \times 100 / XY \quad (A.3.5-8)$$

### A.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 (véase la Figura A.6).

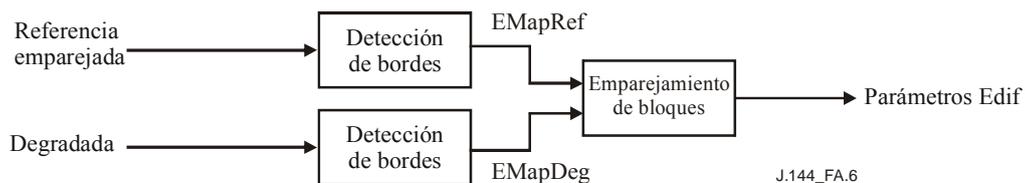


Figura A.6 – Análisis de los bordes

#### A.3.6.1 Detección de los bordes

Se ha utilizado un detector de bordes Canny [A-2] 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  $(x, y)$ :

$$EMap(x, y) = 1 \quad x = 0..X - 1, \quad y = 0..Y - 1 \quad (A.3.6.1-1)$$

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

$$EMap(x, y) = 0 \quad x = 0..X - 1, \quad y = 0..Y - 1 \quad (A.3.6.1-2)$$

### A.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 A.3.6.2-1 a A.3.6.2-5.

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.

$$bref(x, y) = \sum_{i=i1}^{i2} \sum_{j=j1}^{j2} EMapRe f(Nx + X1 + i, My + Y1 + j) \quad x = 0..Bh-1, y = 0..Bv-1 \quad (A.3.6.2-1)$$

$$BDeg(x, y) = \sum_{i=i1}^{i2} \sum_{j=j1}^{j2} EMapDeg(Nx + X1 + i, My + Y1 + j) \quad x = 0..Bh-1, y = 0..Bv-1 \quad (A.3.6.2-2)$$

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

$$i1 = -(N \text{ div } 2) \quad i2 = (N - 1) \text{ div } 2 \quad (A.3.6.2-3)$$

$$j1 = -(M \text{ div } 2) \quad j2 = (M - 1) \text{ div } 2 \quad (A.3.6.2-4)$$

donde el operador "div" representa la división entera.

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

$$EDif = (1/(N \times M \times Bh \times Bv)) \times \left( \sum_{x=0}^{Bh-1} \sum_{y=0}^{Bv-1} (BRe f(x, y) - BDeg(x, y))^Q \right)^{1/Q} \quad (A.3.6.2-5)$$

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

$$N = 4, \quad X1 = 6, \quad Bh = 178, \quad M = 4, \quad Y1 = 10, \quad Bv = 69 \quad Q = 3 \quad (A.3.6.2-6)$$

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

$$N = 4, \quad X1 = 6, \quad Bh = 178, \quad M = 4, \quad Y1 = 10, \quad Bv = 58, \quad Q = 3 \quad (A.3.6.2-7)$$

### A.3.7 Análisis de la PSNR emparejada

Se calcula la relación señal-ruido emparejada para los valores del píxel  $V$  utilizando los vectores de emparejamiento definidos por las ecuaciones A.3.3.3-1 a A.3.3.3-3. Para cada conjunto de vectores emparejados se calcula la medida del error,  $VE$ , mediante la expresión:

$$VE(h, v) = (1/81) \sum_{i=-4}^4 \sum_{j=-4}^4 (DegV(N, Px(h) + i, Py(h) + j) - Re fVField(N + SyncT(h, v), Px(h) + SyncX(h, v) + i, Py(v) + SyncY(h, v) + j))^2 \quad (A.3.7-1)$$

A continuación se calcula la medida de la PSNR segmentaria del campo mediante la expresión:

$$SegVPSNR = (1/Qx \times Qy) \sum_{h=0}^{Qx-1} \sum_{v=0}^{Qy-1} 10,0 \times \log_{10}(255^2 / (VE(h, v) + 1)) \quad (A.3.7-2)$$

#### A.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 A.4-1:

$$AvD(k) = (1/N) \times \sum_{n=0}^{N-1} D(k, n) \quad k = 0..5 \quad (A.4-1)$$

siendo  $N$  el número total de campos de las secuencias probadas y  $D(k, n)$  el 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 expresión:

$$PDMOS = Offset + \sum_{k=0}^5 AvD(k) \times W(k) \quad (A.4-2)$$

Los Cuadros A.2 y A.3 muestran los parámetros del integrador para las secuencias de 625 y 525 respectivamente.

**Cuadro A.2 – Parámetros de integración para el vídeo de radiodifusión de 625**

K	Nombre del parámetro	W
0	TextureDeg	-0,68
1	PySNR(3,3)	-0,57
2	EDif	+58913,294
3	fXPerCent	-0,208
4	MPSNR	-0,928
5	SegVPSNR	-1,529
Offset	+176,486	
N	400	

**Cuadro A.3 – Parámetros de integración para el vídeo de radiodifusión de 525**

K	Nombre del parámetro	W
0	TextureDeg	+0,043
1	PySNR(3,3)	-2,118
2	EDif	+60865,164
3	fXPerCent	-0,361
4	MPSNR	+1,104
5	SegVPSNR	-1,264
Offset	+260,773	
N	480	

## A.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 están soportados en este modelo siendo necesario un módulo de registro independiente para garantizar la correcta alineación de los ficheros de referencia y degradado.

## A.6 Referencias

- [A-1] Material de referencia UIT-T (2004), *Objective perceptual assesment of video quality: Full reference television*.
- [A-2] J. CANNY: A computational approach to edge detection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(6): pp. 679-698, 1986.

## A.7 Datos subjetivos y objetivos

NOTA – Los ficheros de vídeo que figuran a continuación fueron procesados para obtener los resultados indicados. Los nombres del ficheros tienen únicamente carácter informativo. No están a disposición del público. Para más información, véase Ref. [Fase II – Informe final].

*Datos subjetivos y objetivos para 525 líneas*

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
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

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
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
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

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
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

*Datos subjetivos y objetivos para 625 líneas*

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

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
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
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

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
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 B

### Yonsei University/SK Telecom/Radio Research Laboratory

#### Descripción funcional de un modelo de calidad de vídeo con referencia completa

##### B.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.

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

###### B.2.1 PSNR de los bordes (EPSNR, *edge PSNR*)

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 notas 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 cuadrático medio se calcula la PSNR 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.

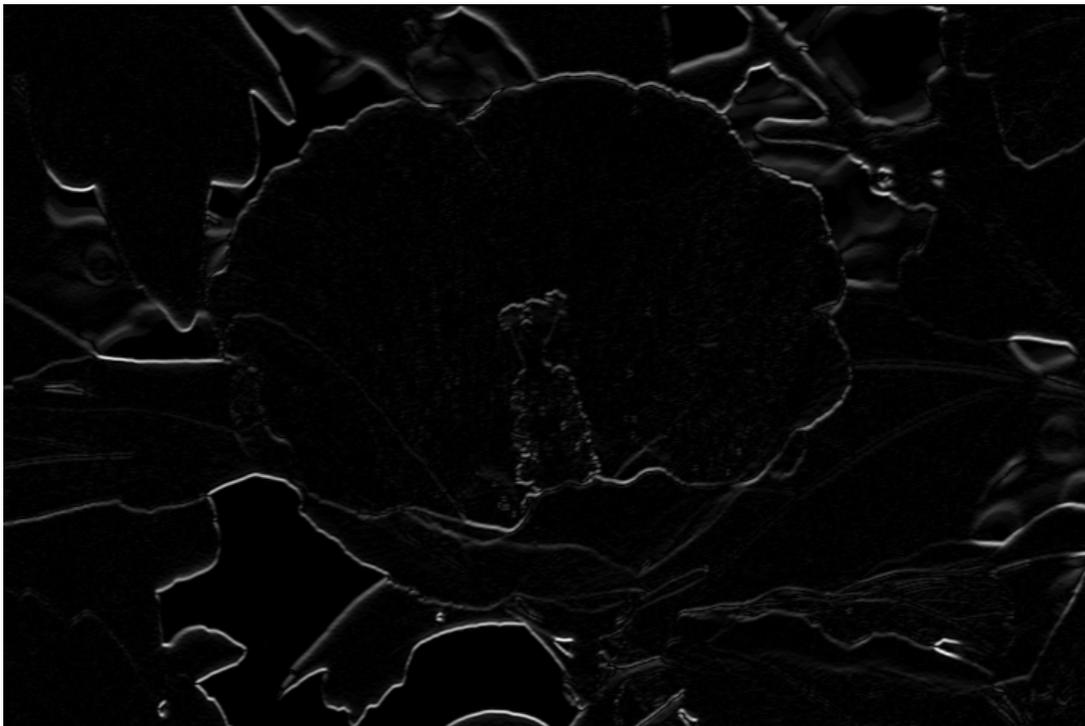
Las Figuras B.1 a B.5 ilustran el procedimiento anterior. La Figura B.1 muestra una imagen fuente. La Figura B.2 muestra la imagen de gradiente horizontal  $g_{horizontal}(m, n)$ , obtenida por aplicación de un operador gradiente horizontal a la imagen fuente de la Figura B.1. La Figura B.3 muestra una imagen de gradiente vertical  $g_{vertical}(m, n)$ , obtenida por aplicación de un operador gradiente vertical a la imagen fuente de la Figura B.1. La Figura B.4 muestra la imagen de gradiente de magnitud (imagen de bordes) mientras que la Figura B.5 muestra la imagen de bordes binarios (imagen máscara) obtenida por aplicación de umbrales a la imagen de gradiente de magnitud de la Figura B.4.



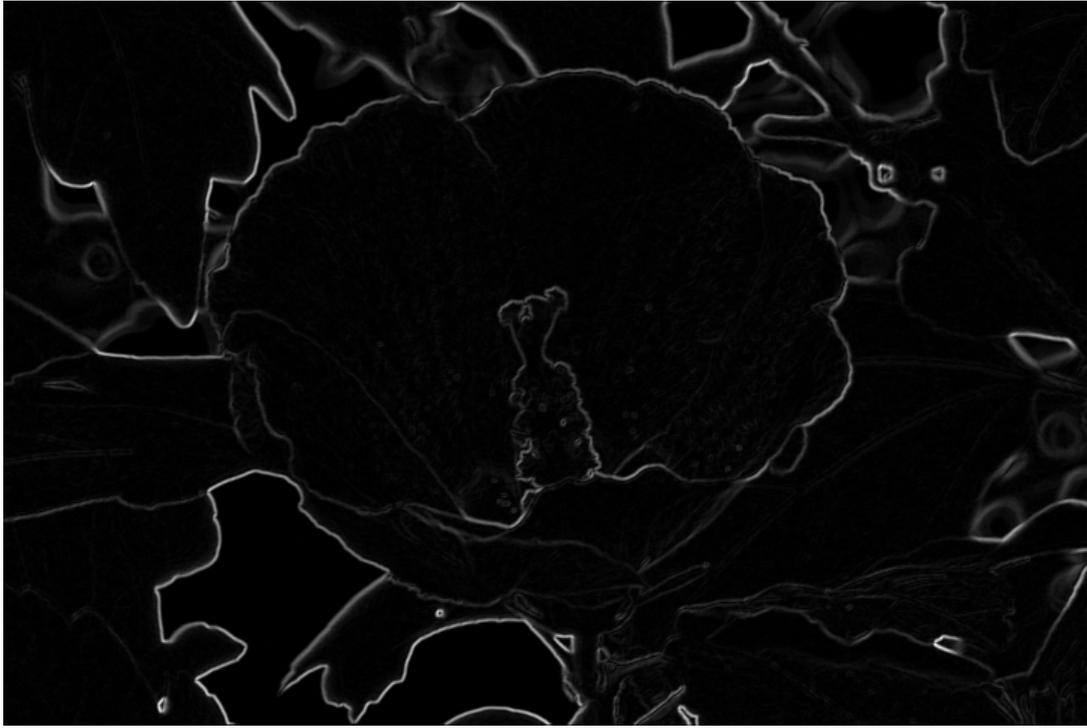
**Figura B.1 – Imagen fuente (imagen original)**



**Figura B.2 – Imagen de gradiente horizontal obtenida por aplicación de un operador gradiente horizontal a la imagen de la Figura B.1**



**Figura B.3 – Imagen de gradiente vertical obtenida por aplicación de un operador gradiente vertical a la imagen fuente de la Figura B.1**

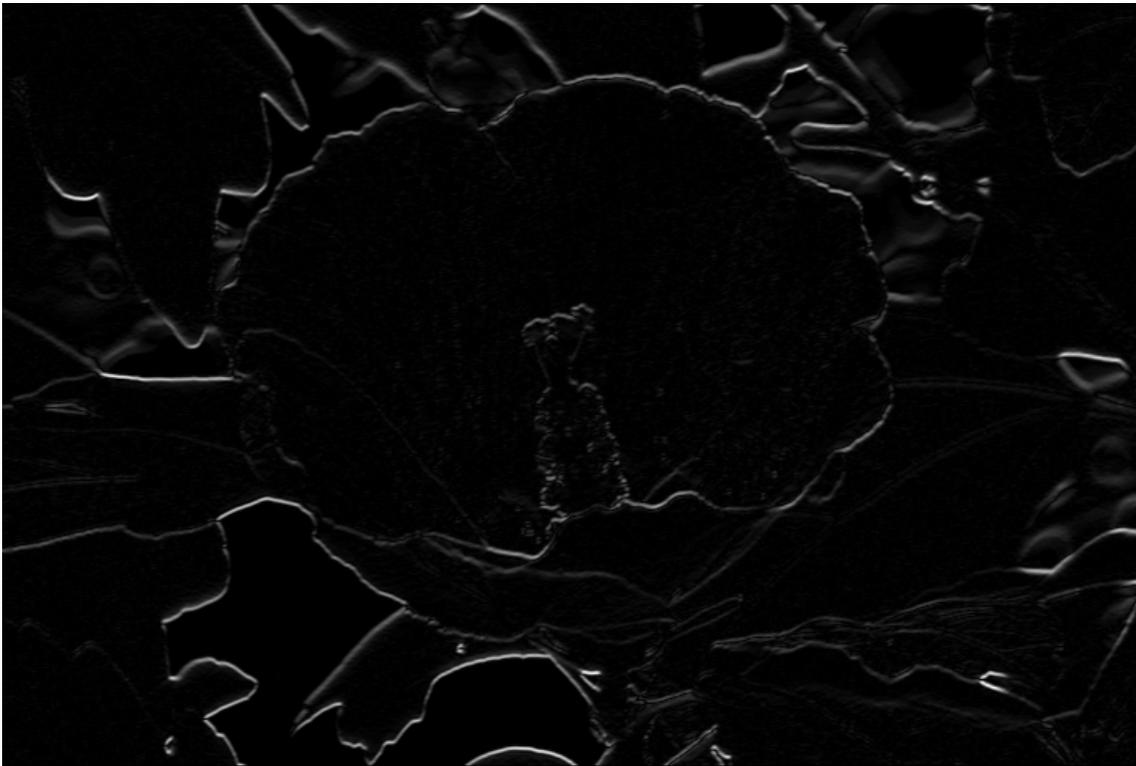


**Figura B.4 – Imagen de gradiente de magnitud**

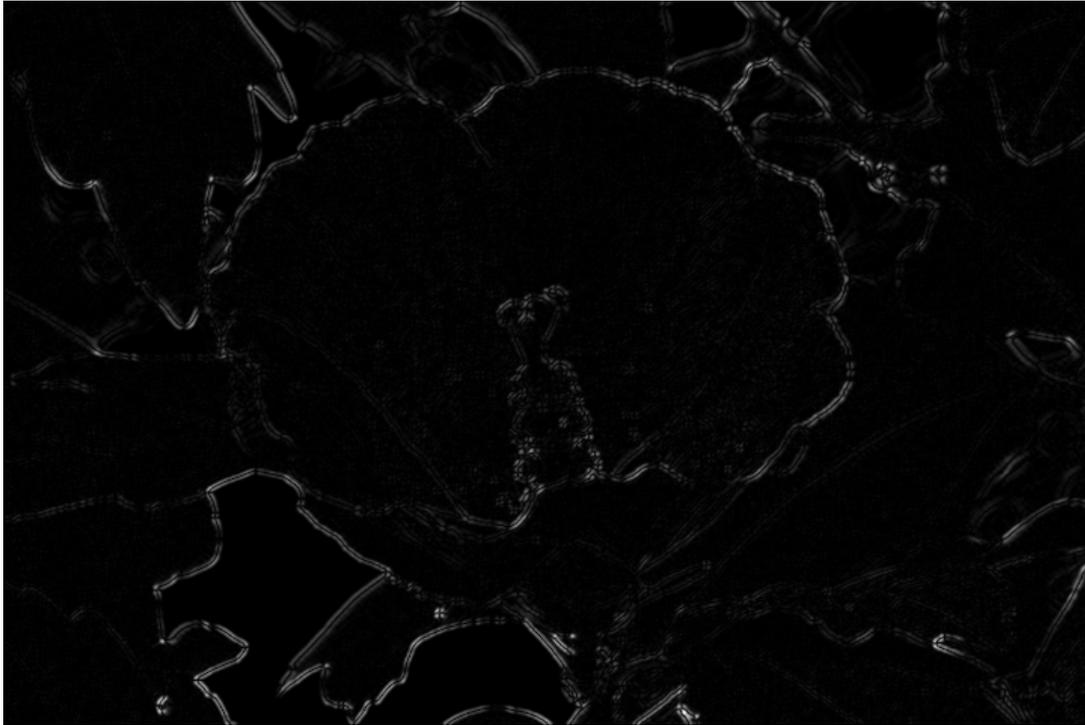


**Figura B.5 – Imagen de bordes binarios (imagen máscara) obtenida por aplicación de umbrales a la imagen de gradiente de magnitud de la Figura B.4**

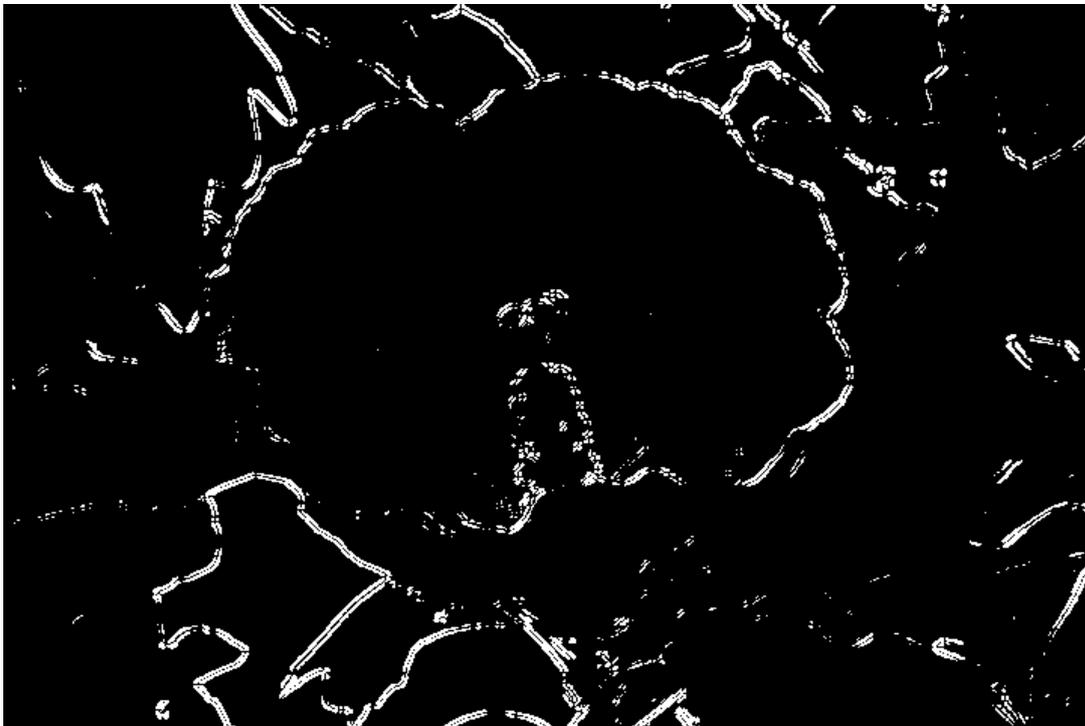
Otra solución consiste en utilizar un procedimiento modificado para determinar las zonas de bordes. Por ejemplo, se puede aplicar en primer lugar un operador gradiente vertical a la imagen fuente, produciendo una imagen de gradiente vertical. A continuación se aplica el operador gradiente horizontal 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 Figuras B.6 a B.9 ilustran el procedimiento modificado. La Figura B.6 muestra una imagen de gradiente vertical  $g_{vertical}(m,n)$ , obtenida por aplicación de un operador gradiente vertical a la imagen fuente de la Figura B.1. La Figura B.7 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 Figura B.6. La Figura B.8 muestra la imagen de bordes binarios (imagen máscara) obtenida por aplicación de umbrales a la imagen modificada por gradientes sucesivos de la Figura B.7.



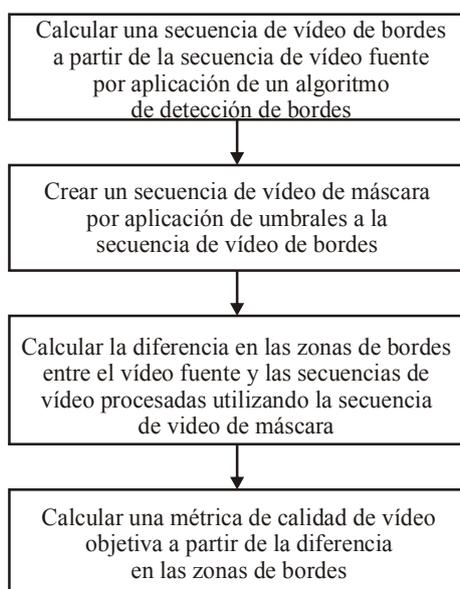
**Figura B.6 – Imagen de gradiente vertical obtenida por aplicación de un operador gradiente vertical a la imagen de la Figura B.1**



**Figura B.7 – 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 Figura B.6**



**Figura B.8 – Imagen de bordes binarios (imagen máscara) obtenida por aplicación de umbrales a la imagen modificada por gradientes sucesivos de la Figura B.7**



J.144\_FB.9

**Figura B.9 – Diagrama de bloques de la EPSNR**

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.

Así pues, en el modelo, se aplica en primer lugar un operador de detección de bordes, produciendo las imágenes de bordes (Figuras B.4 y B.7). A continuación se genera una imagen de máscara (imagen de bordes binarios) por aplicación de umbrales a la imagen de bordes (Figuras B.5 y B.8). 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 Figuras B.5 y B.8 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:

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

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

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

siendo  $Q^l(i,j)$  la  $l$ -ésima imagen de la secuencia de vídeo de bordes y  $t_e$  un umbral. Aunque en la ecuación B-1 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:

$$mse_e = \frac{1}{K} \sum_{l=1}^L se_e^l \quad (B-3)$$

siendo  $L$  el número de imágenes (cuadros o campos) y  $K$  el número total de píxeles de las zonas de bordes. Por último, la PSNR de las zonas de bordes se calculó del siguiente modo:

$$EPSNR = 10 \log_{10} \left( \frac{P^2}{mse_e} \right) \quad (B-4)$$

siendo  $P$  el valor de cresta del píxel. En el modelo, esta PSNR de bordes (EPSNR) se utiliza como nota básica de la calidad objetiva de vídeo. La Figura B.9 muestra el diagrama de bloques para el cálculo de la EPSNR.

## B.2.2 Ajustes posteriores

### B.2.2.1 Deénfasis 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:

$$EPSNR = \begin{cases} EPSNR & \text{si } 0 \leq EPSNR \leq 35 \\ EPSNR \times 0.9 & \text{si } 35 \leq EPSNR \leq 40 \\ EPSNR \times 0.8 & \text{si } EPSNR > 40 \end{cases} \quad (B-5)$$

### B.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:

$$MEPSNR = \begin{cases} EPSNR - 60 \times \left( 0,1225 - \left( \frac{EP_{common}}{EP_{src}} \right)^2 \right) & \text{si } EPSNR < 25 \text{ y } \frac{EP_{common}}{EP_{src}} < 0,35 \text{ y } \frac{EP_{hrc}}{EP_{src}} < 0,13 \\ EPSNR & \text{en los demás casos} \end{cases} \quad (B-6)$$

siendo:

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

$EP_{hrc}$ : el número total de píxeles de bordes en la secuencia de vídeo procesadas

$EP_{common}$ : el 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)

$MEPSNR$ : la EPSNR modificada.

En algunas secuencias de vídeo,  $EP_{src}$  puede ser muy pequeña. En el caso más desfavorable,  $EP_{src}$  puede ser nulo (imagen en blanco o imágenes a muy baja frecuencia), causando un error de división 0. Para evitar estos casos las modificaciones siguientes se recomiendan: Si  $EP_{src}$  es menor que 10 000 píxeles (aproximadamente  $10\,000/240 = 41,7$  píxeles por cuadro para los vídeos de 528 y 8 segundos y aproximadamente  $10\,000/200 = 50$  píxeles por cuadro para los vídeos de 625 y 8 segundos), el usuario puede reducir el umbral  $t_e$  de la ecuación B-2 en 20 hasta que  $EP_{src}$  sea mayor o igual que 10 000 píxeles. Si  $EP_{src}$  es menor que 10 000 píxeles a pesar de haber reducido  $t_e$  a 80, no se utiliza el ajuste posterior de la ecuación B-6. 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 B-6.

### B.2.2.3 Ajuste a escala

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

$$VQM = 1 - MEPSNR \times 0,02 \quad (B-7)$$

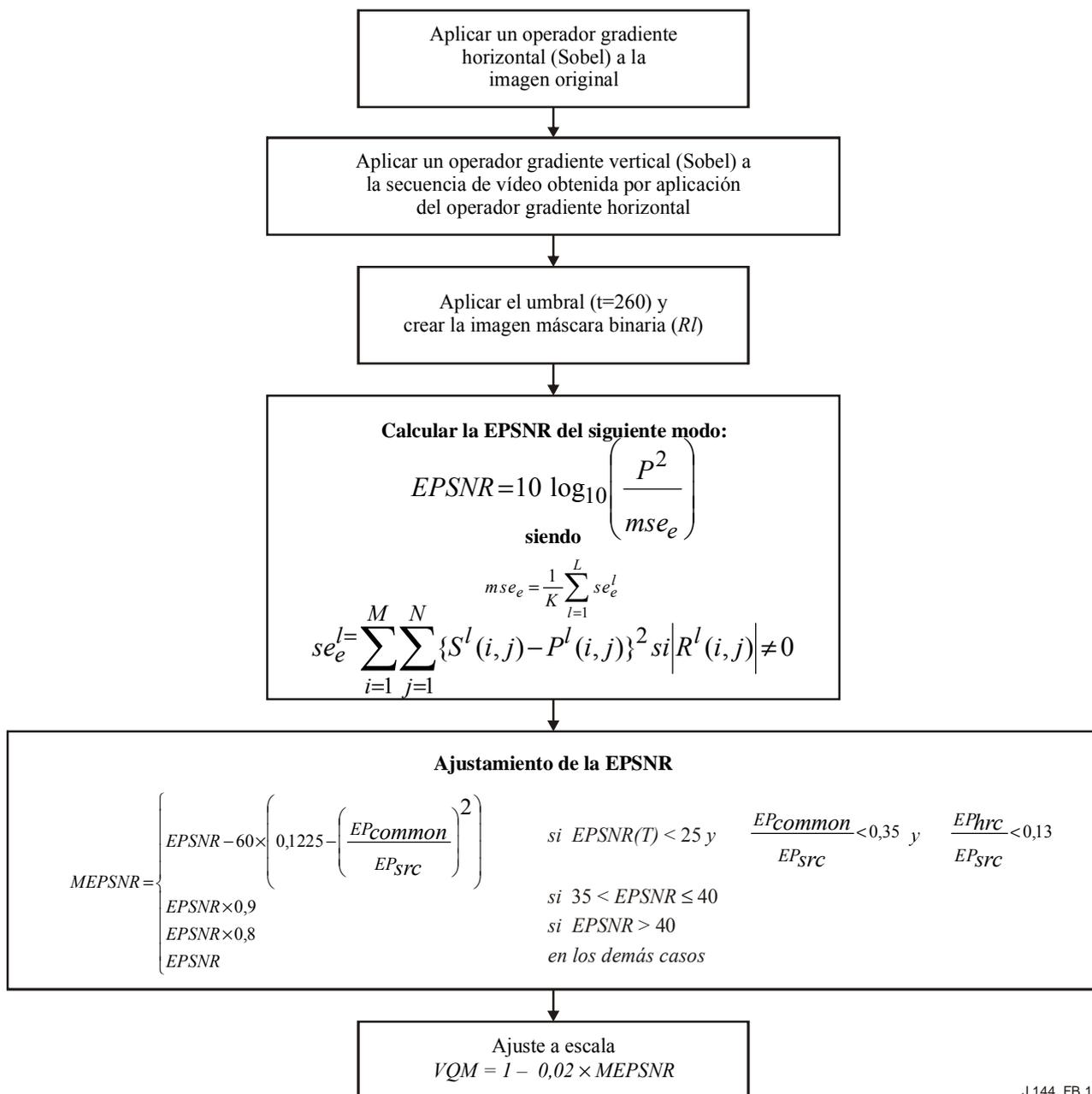
Esta VQM se utiliza como nota objetiva del modelo.

### B.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, como mínimo, para calcular los valores de los subpíxeles.

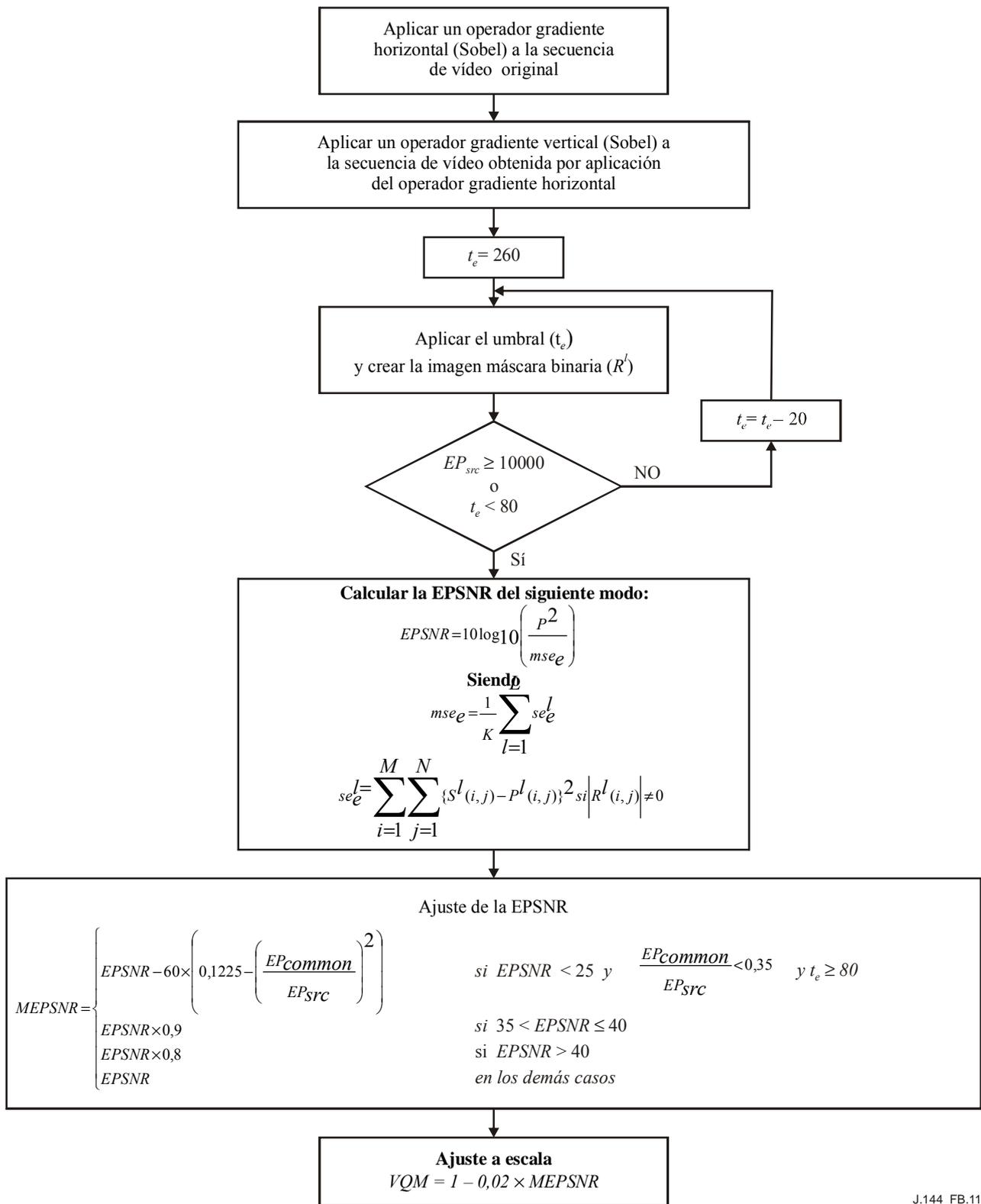
### B.2.4 Diagrama de bloques del modelo

La Figura B.10 muestra el diagrama de bloques completo del modelo. Por otra parte, la Figura B.11 representa un diagrama de bloques modificado que evita el error de división por cero.



J.144\_FB.10

**Figura B.10 – Diagrama de bloques completo del modelo (en este modelo  $P=255$ )**



J.144\_FB.11

**Figura B.11 – Diagrama de bloques modificado para evitar el error de división por cero (en este modelo  $P=255$ )**

### B.3 Registro

#### B.3.1 Registro de vídeo

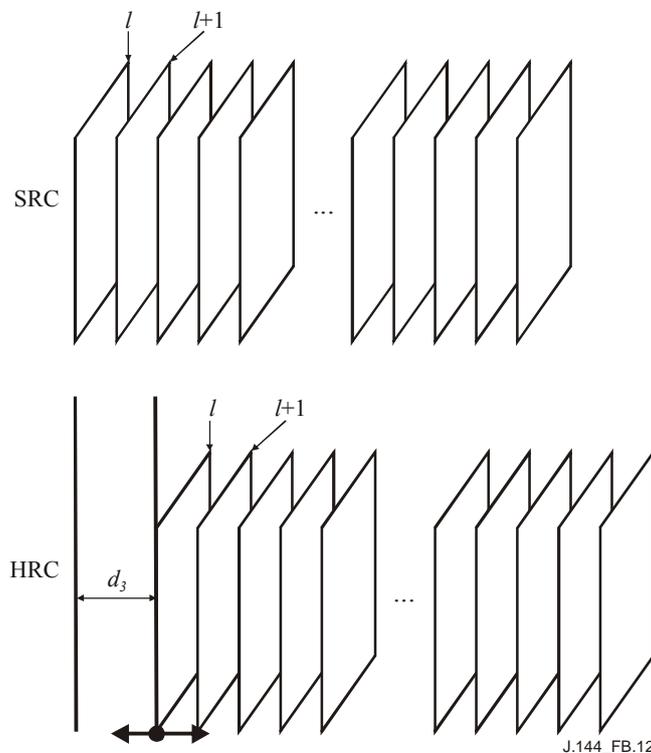
El registro de vídeo es necesario para establecer la mejor concordancia posible entre dos secuencias de vídeo. En la evaluación de la calidad de vídeo, es necesario averiguar cuánto está desplazado el vídeo procesado en las direcciones espacial y temporal. Si la traslación se representa por el vector desplazamiento  $D=[d_1, d_2, d_3]^T$ , el error cuadrático medio entre una secuencia de vídeo fuente y una secuencia de vídeo procesado traslado por  $D$  se calcula del siguiente modo

$$MSE(d_1, d_2, d_3) = \frac{1}{LMN} \sum_l^L \sum_m^M \sum_n^N (U(m, n, l) - V(m + d_1, n + d_2, l + d_3))^2 \quad (\text{B-8})$$

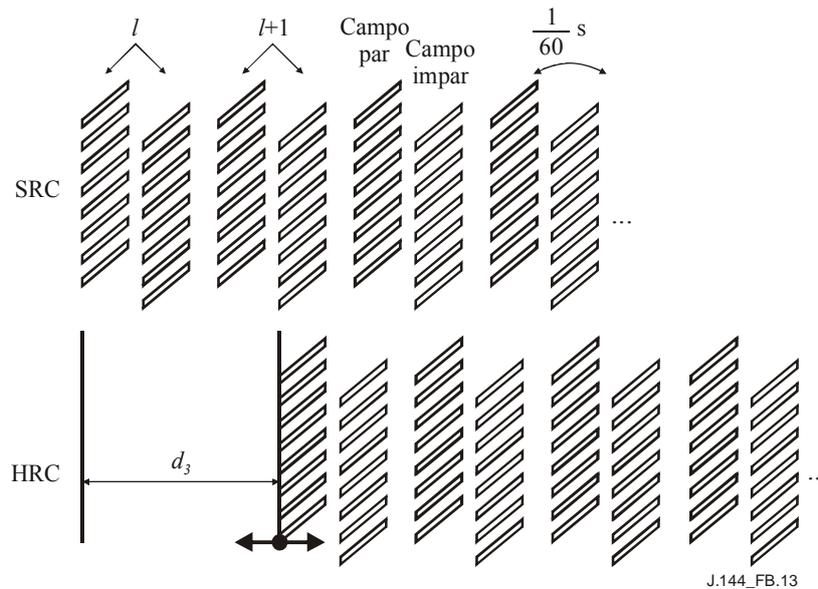
siendo  $U$  y  $V$  las secuencias de vídeo. La estimación del vector desplazamiento que proporciona la mejor concordancia, siendo por tanto la mejor, se obtiene reduciendo al mínimo el  $MSE$ :

$$\hat{D} = \arg \min_{(d_1, d_2, d_3)} MSE(d_1, d_2, d_3) \quad (\text{B-9})$$

La precisión de la componente vertical u horizontal del vector desplazamiento puede ser un número de píxeles entero o fraccionario. Si es fraccionario, hay que utilizar una técnica tal como la interpolación bilineal o la interpolación spline cúbica. Normalmente, la precisión de la componente temporal del vector desplazamiento es un cuadro de vídeo para una secuencia de vídeo explorada progresivamente, como puede verse en la Figura B.12. Para una secuencia de vídeo entrelazado, es un desplazamiento de un campo: 1/50 s para vídeos entrelazados de 50 Hz y 1/60 s para vídeos entrelazados de 60 Hz, como puede verse en la Figura B.13. En el formato entrelazado, es necesario construir un cuadro completo a partir de cada campo a fin de averiguar el desplazamiento espacial de los vídeos entrelazados.



**Figura B.12 – Registro temporal de la secuencia de vídeo progresivo.**  
 $d_3$  se expresa en unidades de cuadros de vídeo



**Figura B.13 – Registro temporal de secuencias de vídeo entrelazado.**  
 $d_3$  se expresa en unidades de campo (1/60 s para las secuencias de vídeo entrelazado de 60 Hz)

### B.3.2 Registro de vídeo basado en la región de interés

Encontrar la mejor concordancia por medio de las ecuaciones B-8 y B-9 suele requerir un tiempo de proceso muy largo. Si se requiere un registro preciso, pueden utilizarse todos los cuadros de la secuencia de vídeo para encontrar el  $MSE$ . No obstante, su proceso llevaría mucho tiempo y no podría efectuarse en tiempo real. Para reducir el tiempo de proceso, en el modelo se selecciona de la secuencia de vídeo un pequeño número de subregiones (regiones de interés). A continuación, el modelo encuentra la mejor concordancia entre las dos secuencias de vídeo calculando el  $MSE$  en las regiones de interés (ROI):

$$MSE_{ROI}(d_1, d_2, d_3) = \frac{1}{K} \sum_{(m,n,l) \in ROI} (U(m, n, l) - V(m + d_1, n + d_2, l + d_3))^2 \quad (B-10)$$

siendo  $K$  el número de píxeles de las ROI.

Se observa que las zonas que cambian con rapidez en las secuencias de vídeo proporcionan una información valiosa para el registro de imagen. Por consiguiente, el modelo localiza estas zonas y las utiliza para el registro de vídeo. Para encontrar escenas que cambien rápidamente, se calcula el error cuadrático medio del cuadro ( $fMSE$ ) del  $l$ -ésimo cuadro de una secuencia de vídeo fuente del siguiente modo:

$$fMSE(l) = \frac{1}{MN} \sum_m^M \sum_n^N (U(m, n, l) - U(m, n, l + 1))^2 \quad (B-11)$$

Tras calcular  $fMSE(l)$  para todos los cuadros de una secuencia de vídeo, el modelo selecciona cinco cuadros de referencia con  $fMSE$  máximos. Estos cuadros pueden considerarse como las escenas que cambian más rápidamente en dirección temporal. Además, el modelo tiene una restricción adicional consistente en que los intervalos entre los cuadros de referencia deben sobrepasar un cierto periodo de tiempo. A tal efecto el modelo divide uniformemente toda la secuencia de vídeo en cinco subsecuencias. De cada subsecuencia, el modelo selecciona un cuadro que es el de mayor  $fMSE$  de la ecuación B-11.

De entre los cinco cuadros seleccionados, el modelo selecciona el cuadro con el mayor MSE de cuadro. A continuación el modelo aplica a la trama una transformada ondícula 2-D y las zonas en las que predominan los coeficientes de alta frecuencia se consideran zonas de cambio rápido en el dominio espacial. La transformada ondícula 2-D se efectúa aplicando por separado transformadas ondícula 1-D en las direcciones horizontal y vertical. La transformada ondícula 1-D se calcula del siguiente modo:

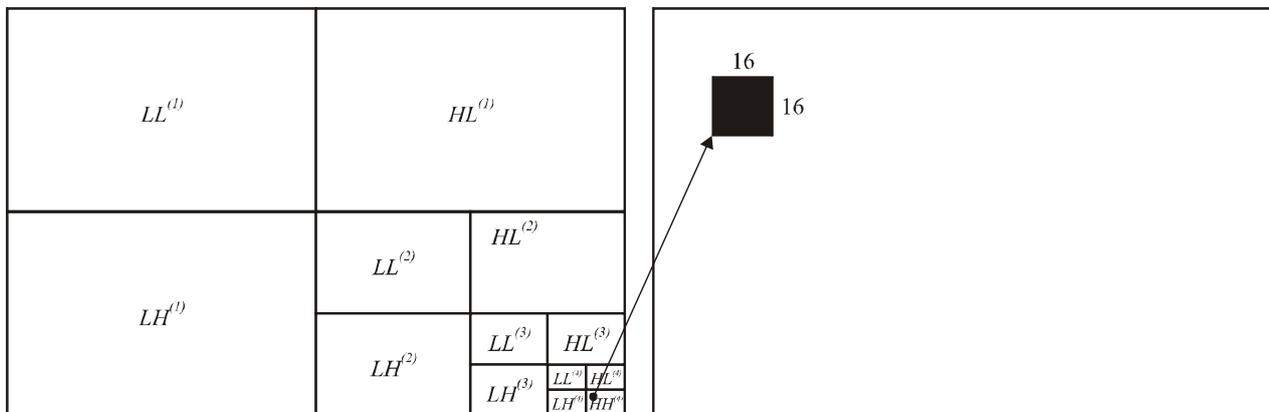
$$y_L^{(1)}[n] = \sum_k x[k]h_0[2n-k] \quad (B-12)$$

$$y_H^{(1)}[n] = \sum_k x[k]h_1[2n-k] \quad (B-13)$$

siendo  $x[n]$  una señal de entrada 1-D y  $h_0[n]$ ,  $h_1[n]$  las respuestas de impulso de los filtros de análisis paso bajo y paso alto (filtros Harr). Los términos  $y_L^{(1)}[n]$ ,  $y_H^{(1)}[n]$  representan las señales de salida de los filtros. El modelo aplica repetidas veces la transformada ondícula a la subbanda de alta frecuencia del siguiente modo:

$$y_H^{(l+1)}[n] = \sum_k y_H^{(l)}h_1[2n-k] \quad (B-14)$$

La Figura B.14 representa la transformada recurrente de la subbanda de alta frecuencia correspondiente a la imagen 2-D. A continuación, el modelo selecciona los doce coeficientes mayores de la subbanda de alta frecuencia y se seleccionan las subregiones correspondientes como ROI, como se puede ver en la Figura B.14. En el modelo, el nivel de descomposición es cuatro. Por consiguiente, el tamaño de las ROI de la imagen original está comprendido entre  $16 \times 16$  y  $(2^4 m, 2^4 n)$ . De este modo, el modelo selecciona doce ROI de  $16 \times 16$  con frecuencias espaciales elevadas.

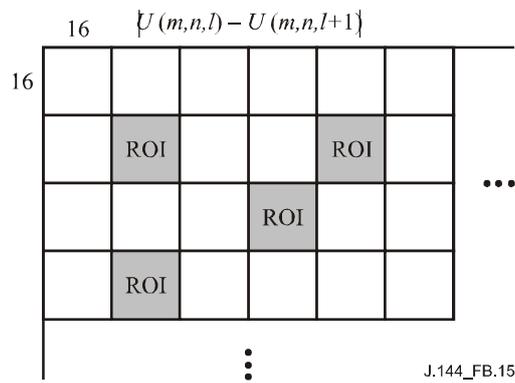


J.144\_FB.14

**Figura B.14 – Transformada ondícula de cuadro y ROI para registro de vídeo rápido**

Para seleccionar ROI que tengan frecuencias temporales elevadas, se divide el cuadro de mayor  $fMSE$  en varios bloques de  $16 \times 16$  (Figura B.15). A continuación, el modelo selecciona los doce bloques de mayor diferencia absoluta de bloques. La diferencia absoluta de bloques ( $ABD$ , *absolute block difference*) se calcula del siguiente modo:

$$ABD = \frac{1}{256} \sum_{(m,n) \in k\text{-ésimo bloque}} |U(m,n,l) - U(m,n,l+1)| \quad (B-15)$$



**Figura B.15 – El modelo selecciona los doce bloques de mayor ABD del cuadro de mayor MSE**

Así pues, el modelo selecciona doce bloques con frecuencias espaciales altas y doce bloques con las mayores *ABD* del cuadro con el mayor MSE de cuadro. Estos 24 bloques se utilizan como ROI. Debe tenerse en cuenta que los bloques ( $24 \times 4=96$  bloques) de los cuatro cuadros restantes, que se encuentran en las mismas posiciones, también se utilizan como ROI. Por consiguiente, se utiliza como ROI un total de 120 bloques de tamaño igual a  $16 \times 16$  y se calcula el mejor vector de desplazamiento mediante la ecuación B-10. Hay que tener también en cuenta que los 24 bloques del cuadro con mayor MSE de cuadro tienen doble peso que los bloques de los cuatro cuadros restantes. Durante el proceso de registro, se efectuó un registro de un cuarto de píxel utilizando la interpolación spline cúbica.

#### **B.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 resulta ser varias veces más rápido que la PSNR convencional, lo que supone una mejora significativa. Así pues, este modelo está perfectamente adaptado a las aplicaciones que requieran la evaluación de la calidad de vídeo en tiempo real.

#### **B.5 Referencias**

- [B-1] Material de referencia UIT-T (2004), *Objective perceptual assessment of video quality: Full reference television*.

**Cuadro B.1 – Matriz VQM de 525 (datos brutos)<sup>2</sup>**

SRC (Image)	HRC=1	HRC=2	HRC=3	HRC=4	HRC=5	HRC=6	HRC=7	HRC=8	HRC=9	HRC=10	HRC=11	HRC=12	HRC=13	HRC=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>2</sup> Tras la presentación del modelo se detectaron errores de registro y del operador. Los datos objetivos 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 de los del Cuadro B.1.

**Cuadro B.2 – Matriz VQM de 625 (datos brutos)<sup>3</sup>**

SRC (Image)	HRC=1	HRC=2	HRC=3	HRC=4	HRC=5	HRC=6	HRC=7	HRC=8	HRC=9	HRC=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>3</sup> Tras la presentación del modelo se detectaron errores de registro y del operador. Los datos objetivos 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 de los del Cuadro B.2.

**Cuadro B.3 – Matriz VQM de 525 (datos ajustados a escala)<sup>4</sup>**

SRC(Image)	HRC=1		HRC=2		HRC=3		HRC=4		HRC=5		HRC=6		HRC=7		HRC=8		HRC=9		HRC=10		HRC=11		HRC=12		HRC=13		HRC=14	
1	1	0,727	4	0,490	7	0,467	10	0,304																				
2	2	0,324	5	0,224	8	0,162	11	0,195																				
3	3	0,549	6	0,359	9	0,192	12	0,153																				
4									13	0,715	17	0,588	21	0,546	25	0,509												
5									14	0,523	18	0,418	22	0,344	26	0,289												
6									15	0,665	19	0,404	23	0,341	27	0,292												
7									16	0,386	20	0,298	24	0,239	28	0,199												
8																	29	0,823	35	0,783	41	0,788	47	0,537	53	0,481	59	0,512
9																	30	0,854	36	0,550	42	0,774	48	0,435	54	0,377	60	0,383
10																	31	0,539	37	0,354	43	0,520	49	0,206	55	0,156	61	0,241
11																	32	0,634	38	0,662	44	0,669	50	0,400	56	0,393	62	0,442
12																	33	0,580	39	0,515	45	0,581	51	0,390	57	0,332	63	0,353
13																	34	0,542	40	0,568	46	0,476	52	0,273	58	0,247	64	0,293

<sup>4</sup> Tras la presentación del modelo se detectaron errores de registro y del operador. Los datos objetivos 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 de los del Cuadro B.3.

**Cuadro B.4 – Matriz VQM de 625 (datos ajustados a escala)<sup>5</sup>**

SRC(Image)	HRC=1		HRC=2		HRC=3		HRC=4		HRC=5		HRC=6		HRC=7		HRC=8		HRC=9		HRC=10	
1			4	0,625	10	0,429	16	0,204			29	0,164			42	0,169			52	0,082
2			5	0,467	11	0,454	17	0,202			30	0,170			43	0,177			53	0,062
3			6	0,542	12	0,538	18	0,327			31	0,275			44	0,272			54	0,134
4			7	0,605	13	0,686	19	0,502			32	0,475			45	0,414			55	0,255
5			8	0,609	14	0,641	20	0,335			33	0,312			46	0,185			56	0,091
6			9	0,586	15	0,395	21	0,284			34	0,248			47	0,226			57	0,143
7							22	0,551			35	0,496					48	0,430	58	0,339
8							23	0,371			36	0,335					49	0,283	59	0,217
9							24	0,356			37	0,298					50	0,243	60	0,168
10							25	0,623			38	0,590					51	0,549	61	0,455
11	1	0,741							26	0,592			39	0,359					62	0,290
12	2	0,672							27	0,499			40	0,482					63	0,322
13	3	0,698							28	0,406			41	0,279					64	0,179

<sup>5</sup> Tras la presentación del modelo se detectaron errores de registro y del operador. Los datos objetivos 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 de los del Cuadro B.4.

## Anexo C

### Telecommunications Research and Development Center (CPqD)

#### Descripción técnica de la evaluación de imágenes basada en segmentación (IES)

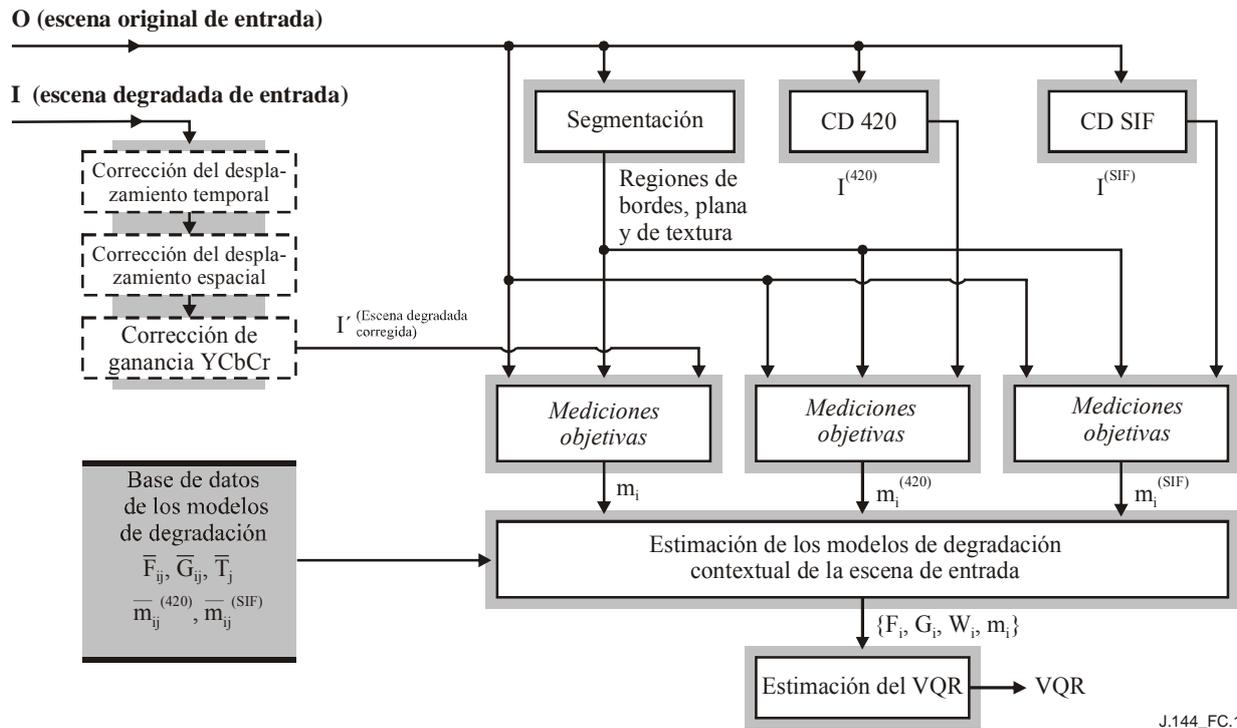
##### C.1 Introducción

En este anexo se presenta una descripción completa del algoritmo de evaluación de imágenes basada en la segmentación (IES, *image evaluation based on segmentation*) propuesto por el Telecommunications Research and Development Center (CPqD). El algoritmo CPqD-IES es una metodología de la evaluación de la calidad de vídeo que utiliza mediciones objetivas de degradación, computadas en regiones planas, de borde y de textura resultantes de un proceso de segmentación de la imagen. Proporciona predicciones sobre el juicio de un observador humano [nota media de opinión (MOS, *mean opinion score*)], lo que constituye una alternativa efectiva y eficiente a las costosas metodologías de pruebas subjetivas de larga duración que se utilizan actualmente [C-1] y [C-2].

El prototipo del programa informático CPqD-IES utilizó en las pruebas de la fase II del VQEG [C-3] y los datos objetivos sin elaborar ajustados a escala obtenidos se representan en los Cuadros C.2 y C.3.

##### C.2 Descripción general del algoritmo CPqD-IES

En la Figura C.1 se representa un esquema del algoritmo CPqD-IES para escenas naturales. Cada escena natural se representa mediante una escena (de referencia) original,  $O$ , y una escena degradada,  $I$ , resultante de aplicar a  $O$  una operación de códec. Se aplican a  $I$  correcciones de desplazamiento y ganancia para crear una escena degradada corregida  $I'$ , de modo que cada cuadro  $f$  de  $I'$  corresponda al cuadro de referencia  $f$  de  $O$ , para  $f = 1, 2, \dots, n$  (cláusula C.3).



J.144\_FC.1

**Figura C.1 – Esquema general del algoritmo CPqD-IES**

Las escenas de entrada  $I$  y  $O$  al algoritmo CPqD-IES tienen el formato YCbCr4:2:2 de la Rec. UIT-R BT.601-5 [C-4].

La componente Y de cada cuadro  $f$  de  $O$  se segmenta en tres categorías: región de textura, región de borde y región plana (cláusula C.4). Una medición objetiva se calcula en base a la diferencia entre los cuadros correspondientes de  $O$  e  $I'$ , para cada uno de estos contextos y para cada componente de imagen Y, Cb y Cr, formando un conjunto de 9 mediciones objetivas  $\{m_1, m_2, \dots, m_9\}$  para cada cuadro  $f$  (cláusula C.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:

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

siendo  $F_i$  y  $G_i$  dos parámetros calculados a partir de la base de datos de los modelos de degradación (cláusula C.6), de los atributos espacial  $S$  y temporal  $T$  (cláusula C.5), y de las mediciones objetivas  $m_i^{(420)}$  y  $m_i^{(SIF)}$  del cuadro  $f$ , resultantes de las operaciones códec CD420 y CDSIF (aplicadas a  $O$  (cláusula C.7)). Los códecs de degradación de referencia CD420 (codificador/decodificador MPEG-2 4:2:0) y CDSIF (codificador/decodificador MPEG-1 SIF), se basan totalmente en las rutinas extraídas directamente de MPEG2 [C-5] y MPEG1 [C-6], que se pueden encontrar en [www.mpeg.org/MPEG/MSSG](http://www.mpeg.org/MPEG/MSSG). En la actual implementación del algoritmo CPqD-IES, estas rutinas funcionan en modo intra utilizando un paso de cuantificación fijo de 16. Es importante observar que CD420 y CDSIF no introducen desplazamiento ni diferencia de ganancia con respecto a  $O$ .

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:

$$VQR_f = \sum_{i=1}^9 W_i \cdot L_i \quad (\text{C-2})$$

siendo  $W_i$  el peso del nivel de degradación  $L_i$  correspondiente a esta escena natural en particular, que se calcula con arreglo a la cláusula C.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 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 entre  $VQR_{f-1}, VQR_f$ , y  $VQR_{f+1}$ . 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 índice final de calidad de vídeo  $VQR$  es el promedio de los valores  $VQR'_f$ .

$$VQR = \frac{1}{n-2} \cdot \sum_{f=1}^{n-2} VQR'_f \quad (C-3)$$

Las ecuaciones C-1 y C-2, 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 Figura C.1.

### C.3 Corrección del desplazamiento y de la ganancia

#### C.3.1 Desplazamiento temporal

El desplazamiento temporal  $dt$  es un entero comprendido entre  $-2$  y  $2$  cuadros. Los valores positivos de  $dt$  significan que la escena degradada  $I$  está retrasada con respecto a la escena original  $O$ , mientras que los valores negativos de  $dt$  significan lo contrario: la escena original  $O$  está retrasada con respecto a la escena degradada  $I$ . Las escenas de entrada con desplazamientos temporales fuera de este rango no son consideradas por el modelo. Sea  $I_{dt}$  la escena  $I$  degradada con un desplazamiento de  $dt$  cuadros. Se calcula un coeficiente de desemejanza entre  $O$  y cada escena desplazada  $I_{dt}$ . El desplazamiento cuyo coeficiente de desemejanza sea el menor se utiliza como desplazamiento temporal y, en el siguiente cálculo, la  $I_{dt}$  de salida es por consiguiente la escena degradada  $I$  desplazada 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 ella:

- 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  $\xi_T$  (coeficiente de desemejanza entre  $O$  e  $I_{dt}$ ).

siendo:

$S_b$  = magnitud del gradiente de Sobel [7] 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  $S_{b'}$ .

### C.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  $d_y = -6, -5, \dots, 6$ . Los valores de desplazamiento negativos significan que un cuadro de la escena degradada  $I$  queda desplazado con respecto a un cuadro de la escena original  $O$  hacia la izquierda y hacia arriba. Los valores de desplazamiento positivos significan que un cuadro de la escena degradada  $I$  queda desplazado con respecto a un cuadro de la escena original  $O$ , hacia la derecha y hacia abajo. El modelo no considera las escenas de entrada con desplazamientos espaciales fuera de este rango.

Sea  $I_{dx, dy}$  la escena degradada  $I_{dt}$  con todos sus cuadros desplazados ( $d_x, d_y$ ) píxeles. Se calcula un coeficiente de desemejanza entre la escena original  $O$  y la escena degradada  $I_{dx, dy}$ . El desplazamiento espacial que produce la menor desemejanza se utiliza como desplazamiento espacial y la  $I_{dx, dy}$  de salida es la escena  $I_{dt}$  desplazada en esta magnitud que se utiliza para la corrección de ganancia.

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:

$w \times h$	la porción de imagen cuyo tamaño es $w$ columnas por $h$ líneas que se presenta simultáneamente en un cuadro $f$ de la escena original $O$ y en el cuadro correspondiente de la escena degradada $I_{dx, dy}$ ;
$Y(x, y), Cb(x, y), Cr(x, y)$	los valores de las componentes de imagen del cuadro $f$ de $O$ para el píxel $(x, y)$ ;
$Y'(x + dx, y + dy),$ $Cb'(x + dx, y + dy),$ $Cr'(x + dx, y + dy)$	los valores de las componentes de imagen del cuadro $f$ ; $I_{dx, dy}$ para el píxel $(x + dx, y + dy)$ .

### C.3.3 Ganancia

La ganancia de amplitud entre  $O$  e  $I_{dx, dy}$  se calcula para cada componente de imagen  $Y, Cb$  y  $Cr$ , 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 [C-7] cuyo núcleo sea

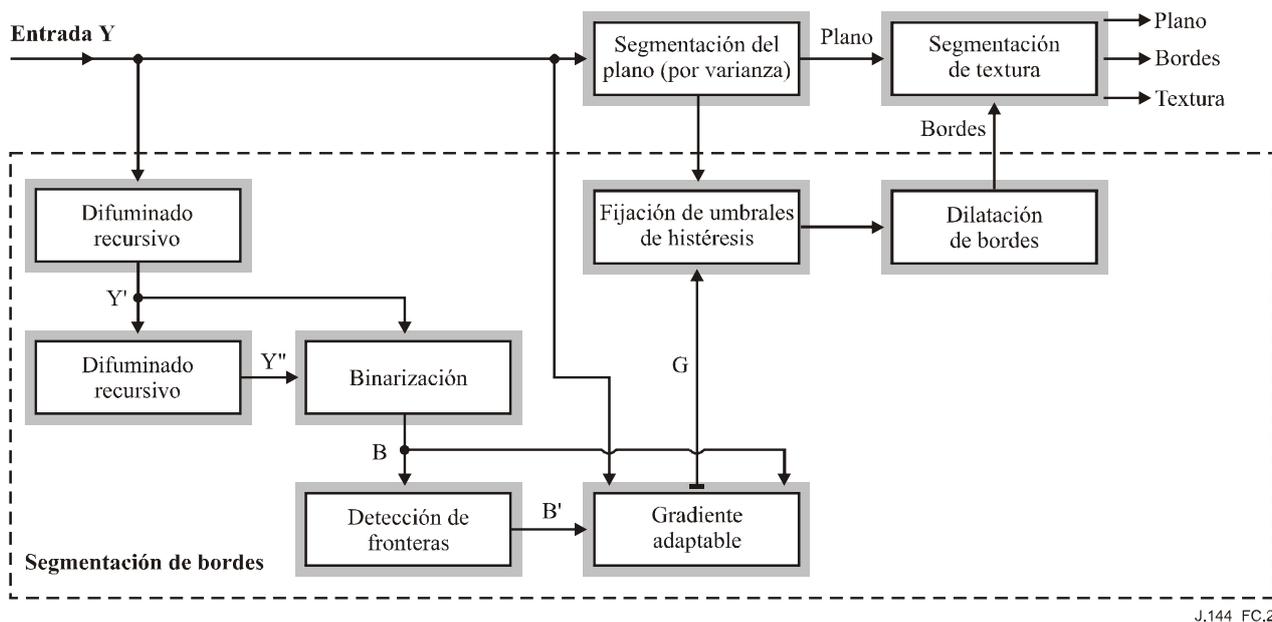
$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}$$

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 C.3.2).

#### C.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 Figura C.2).

La segmentación se computa para cada cuadro de la componente  $Y$  de la escena original de entrada  $O$ . Las regiones segmentadas correspondientes a las componentes  $C_b$  y  $C_r$ , se obtienen reduciendo en un factor de 2 la densidad de muestreo de las posiciones de los píxeles de la componente  $Y$  en sentido horizontal.



J.144\_FC.2

Figura C.2/J.144 – Diagrama de bloques del proceso de segmentación

##### C.4.1 Regiones planas

La varianza del brillo de cada píxel de  $Y$  se calcula en un entorno de  $5 \times 5$  píxeles a su alrededor. Se fijan umbrales para la varianza de la imagen tales que los píxeles cuya 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 por medio de un elemento de estructuración circular de

11 píxeles de diámetro [C-7]. Esta operación corresponde a la aplicación de la ecuación siguiente a una imagen binaria de entrada  $A$  para crear una imagen binaria dilatada  $A'$ .

$$A'(x, y) = \max\{A(x', y')\} \text{ para todos los píxeles } (x', y') \in N_{81}(x, y) \quad (\text{C-4})$$

siendo  $N_{81}(x, y)$  el conjunto de los 81 píxeles más próximos al píxel  $(x, y)$ .

### C.4.2 Región de bordes

Se aplica un filtrado recurrente a  $Y$ , creando una primera imagen difuminada  $Y'$  que se aplica a continuación a  $Y'$  creando una segunda imagen difuminada  $Y''$ . Cada filtrado recurrente 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)  $Y(x, y - 1) \leftarrow Y(x, y) + 0,7.[Y(x, y - 1) - Y(x, y)]$
- 13) Guardar la imagen  $Y$  en la imagen  $Y'$

siendo:

$$\begin{aligned} Y(x, y) &= \text{brillo del píxel } (x, y) \\ h &= \text{número de líneas de } Y \\ w &= \text{número de columnas de } Y. \end{aligned}$$

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

$$B(x, y) = \begin{cases} 1, & \text{si } Y'(x, y) \geq Y''(x, y), \\ 0, & \text{en los demás casos} \end{cases} \quad (\text{C-5})$$

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

$$B'(x, y) = \begin{cases} 1, & \text{si } B(x, y) = 1 \text{ y } B(x', y') = 0 \text{ para cualquier píxel } (x', y') \in N_9(x, y) \\ 0, & \text{en los demás casos} \end{cases} \quad (\text{C-6})$$

siendo  $N_9(x, y)$  el conjunto de los 9 píxeles más próximos a  $(x, y)$ .

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

$$G(x, y) = \begin{cases} |\mu_1 - \mu_0|, & \text{si } B'(x, y) = 1, \\ 0, & \text{en los demás casos,} \end{cases} \quad (\text{C-7})$$

siendo:

$\mu_1 =$  valor medio de  $Y(x', y')$ , para todo  $(x', y') \in N_9(x, y)$  de modo que  $B(x', y') = 1$

$\mu_0 =$  valor medio de  $Y(x', y')$ , para todo  $(x', y') \in N_9(x, y)$  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 [C-8] a  $G$  restringido a los píxeles clasificados en la cláusula C.4.1 como pertenecientes a la región plana. El umbral inferior es 30 y el superior 40. El algoritmo identifica primero como bordes todos los píxeles  $(x, y)$  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 de 5 píxeles de diámetro, es decir, utilizando el conjunto  $N_{13}(x, y)$  de los 13 píxeles más próximos al píxel  $(x, y)$ , análogamente a la ecuación C-4, ignorando la restricción a la región plana. En esta dilatación, los píxeles con valor 1 se clasifican como pertenecientes a la región de bordes.

### C.4.3 Región 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.

### C.5 Medición objetiva

Sea  $S_b$  la imagen de la magnitud del gradiente de Sobel [C-7] calculado para una componente determinada ( $Y$ ,  $Cb$  o  $Cr$ ) de un cuadro determinado  $f$  de la escena original  $O$ , y  $S'_b$ , la imagen de la magnitud del gradiente de Sobel de la misma componente del cuadro  $f$  de la escena degradada  $I'$ . 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) 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  $f$ ,  $f = 1, 2, \dots, n$ , 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  $\{m_1^{(SIF)}, m_2^{(SIF)}, \dots, m_9^{(SIF)}\}$  para el cuadro  $f$  con respecto a las operaciones de los códecs de referencia MPEG-2 4:2:0 y MPEG-1 SIF sobre  $O$  (Figura C.1). Estas mediciones se utilizan como referencia junto con los atributos espacial  $S$  y temporal  $T$  para determinar el modelo de degradación contextual para  $I'$  (cláusula C.7). El atributo temporal  $T$  es el valor medio de la diferencia absoluta en píxeles entre las segmentaciones de los cuadros  $f$  y  $f-1$ , normalizado a  $[0, 1]$ . El atributo espacial  $S$  se define como la relación  $m_7^{(SIF)}/m_7^{(420)}$  normalizada a  $[0, 1]$ , siendo  $m_7^{(SIF)}$  y  $m_7^{(420)}$  las correspondientes ASD para la región plana de la componente  $Y$  del cuadro  $f$ .

### C.6 Base de datos de los modelos de degradación

El algoritmo CPqD-IES utiliza una base de datos de modelos de degradación para escenas distintas a la escena de referencia  $O$  a fin de estimar el índice de calidad de vídeo de  $I'$ . 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  $\{\bar{m}_{1,j}^{(420)}, \bar{m}_{2,j}^{(420)}, \dots, \bar{m}_{9,j}^{(420)}\}$ , y  $\{\bar{m}_{1,j}^{(SIF)}, \bar{m}_{2,j}^{(SIF)}, \dots, \bar{m}_{9,j}^{(SIF)}\}$  para los cuadros de cada escena  $j$ ,  $j = 1, 2, \dots, 12$ . Los valores de  $\bar{T}_j = \{27,01; 25,33; 45,54; 36,40; 32,02; 12,63; 28,38; 10,19; 0,01; 7,26; 7,60; 14,27\}$  se calcularon

como promedio de los atributos temporales, calculados como se describe en cláusula C.5, sobre los cuadros de cada escena  $j$ .

Toda 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 dentro del intervalo [0%, 100%].

Cada medición objetiva  $\bar{m}_{i,j}$ , siendo  $i = 1, 2, \dots, 9$  y  $j = 1, 2, \dots, 12$ , extraída de las escenas de la base de datos (original y degradada) se relaciona con un nivel de degradación contextual  $\bar{L}_{i,j}$ , de acuerdo con la ecuación C.1. Los valores de  $F_{ij}$  y de  $G_{ij}$  de la ecuación C-1 se calcularon para cada escena  $j$  minimizando el error cuadrático medio  $E\left[\left(\bar{SL}_j - \bar{L}_{i,j}\right)^2\right]$ .

Por último, se definió la base de datos de los modelos de degradación con cinco conjuntos de parámetros  $\bar{F}_{i,j}, \bar{G}_{i,j}, \bar{T}_j, \bar{m}_{i,j}^{(420)}, \bar{m}_{i,j}^{(SIF)}$ ,  $i = 1, 2, \dots, 9$  para cada escena  $j, j = 1, 2, \dots, 12$ .

El Cuadro C.1 contiene los valores de  $\bar{F}_{i,j}, \bar{G}_{i,j}, \bar{m}_{i,j}^{(420)}, \bar{m}_{i,j}^{(SIF)}$ , siendo Y, Cb y Cr las componentes de un cuadro y P, E y T los sufijos que indican regiones planas, regiones de bordes y regiones de textura, respectivamente.

**Cuadro C.1 – Mediciones de degradación para 12 escenas de la base de datos:  $\bar{m}_{i,j}^{(420)}, \bar{m}_{i,j}^{(SIF)}, \bar{F}_{i,j}, \bar{G}_{i,j}$**

Escena j	$\bar{m}_{1,j}^{(420)}$ YP	$\bar{m}_{2,j}^{(420)}$ CbP	$\bar{m}_{3,j}^{(420)}$ CrP	$\bar{m}_{4,j}^{(420)}$ YE	$\bar{m}_{5,j}^{(420)}$ CbE	$\bar{m}_{6,j}^{(420)}$ CrE	$\bar{m}_{7,j}^{(420)}$ YT	$\bar{m}_{8,j}^{(420)}$ CbT	$\bar{m}_{9,j}^{(420)}$ CrT
1	10,89	7,08	7,41	22,73	24,21	24,68	22,93	20,94	19,25
2	11,69	5,82	5,12	20,06	15,80	12,22	21,55	12,41	9,76
3	8,14	5,36	4,32	13,77	12,54	11,21	14,43	10,89	10,43
4	14,18	6,04	5,44	21,89	15,57	12,19	21,50	12,55	10,80
5	6,87	5,44	4,50	19,42	17,24	14,40	20,36	19,18	16,98
6	8,96	4,31	4,17	16,67	8,08	10,56	15,58	6,38	6,15
7	14,25	8,10	6,99	22,69	17,65	18,62	21,66	16,80	16,57
8	7,06	3,36	3,92	21,40	17,45	22,31	22,44	30,12	22,21
9	8,80	6,61	7,19	20,65	17,24	13,81	21,02	15,94	12,51
10	16,04	15,92	10,28	19,58	20,93	12,59	19,57	25,38	14,39
11	6,70	4,41	4,98	17,48	11,97	13,50	17,42	15,27	16,44
12	12,10	7,09	8,14	21,49	25,22	22,38	21,58	19,83	19,18

**Cuadro C.1 – Mediciones de degradación para 12 escenas  
de la base de datos:  $\bar{m}_{i,j}^{(420)}$ ,  $\bar{m}_{i,j}^{(SIF)}$ ,  $\bar{F}_{i,j}$ ,  $\bar{G}_{i,j}$**

<b>Escena j</b>	$\bar{m}_{1,j}^{(SIF)}$ <b>YP</b>	$\bar{m}_{2,j}^{(SIF)}$ <b>CbP</b>	$\bar{m}_{3,j}^{(SIF)}$ <b>CrP</b>	$\bar{m}_{4,j}^{(SIF)}$ <b>YE</b>	$\bar{m}_{5,j}^{(SIF)}$ <b>CbE</b>	$\bar{m}_{6,j}^{(SIF)}$ <b>CrE</b>	$\bar{m}_{7,j}^{(SIF)}$ <b>YT</b>	$\bar{m}_{8,j}^{(SIF)}$ <b>CbT</b>	$\bar{m}_{9,j}^{(SIF)}$ <b>CrT</b>
1	21,65	9,88	10,97	83,35	35,03	35,40	72,68	28,46	27,22
2	17,85	6,97	5,90	68,56	20,50	15,08	53,11	15,88	11,67
3	12,57	7,05	5,20	32,24	17,38	14,42	32,80	13,54	12,48
4	21,77	7,17	6,20	75,52	20,10	15,20	61,95	15,88	13,17
5	11,79	6,07	5,24	88,69	22,84	19,79	84,11	25,62	23,44
6	11,85	4,63	4,49	43,21	9,83	13,09	26,98	7,04	6,79
7	21,35	8,62	8,13	110,79	20,89	23,41	88,45	20,36	21,23
8	9,97	3,82	4,48	70,29	22,30	30,25	72,19	40,05	28,45
9	18,27	7,98	8,20	61,46	21,79	16,82	54,04	19,81	15,33
10	27,18	22,33	12,85	42,67	30,69	15,35	39,09	36,27	19,24
11	8,66	4,70	5,73	51,38	14,96	18,01	42,91	19,17	21,01
12	13,99	10,33	11,11	76,35	43,82	35,30	62,17	31,69	31,35
<b>Escena j</b>	$\bar{F}_{1,j}$ <b>YP</b>	$\bar{F}_{2,j}$ <b>CbP</b>	$\bar{F}_{3,j}$ <b>CrP</b>	$\bar{F}_{4,j}$ <b>YE</b>	$\bar{F}_{5,j}$ <b>CbE</b>	$\bar{F}_{6,j}$ <b>CrE</b>	$\bar{F}_{7,j}$ <b>YT</b>	$\bar{F}_{8,j}$ <b>CbT</b>	$\bar{F}_{9,j}$ <b>CrT</b>
1	19,67	9,60	10,34	64,69	35,11	34,99	62,74	30,48	28,92
2	16,80	7,02	5,80	50,94	20,85	15,31	50,94	20,85	15,31
3	16,25	8,43	5,97	48,49	19,54	15,89	50,17	15,57	13,89
4	20,59	7,04	6,01	52,95	20,70	15,58	49,29	16,91	13,90
5	10,64	6,03	5,39	58,91	23,39	19,79	60,79	27,51	24,66
6	11,01	4,48	4,36	29,93	9,27	12,43	24,51	6,84	6,58
7	20,56	8,49	7,91	69,41	20,68	23,91	60,70	20,06	22,14
8	10,18	3,88	4,52	58,20	22,37	30,64	61,92	43,33	31,30
9	24,49	8,92	9,02	70,80	24,17	19,21	63,14	23,06	18,00
10	22,55	20,91	12,45	32,29	29,62	15,01	32,45	36,43	19,03
11	8,03	4,68	5,61	32,73	14,48	16,94	33,55	19,15	20,70
12	13,04	9,30	10,01	44,95	40,64	32,98	45,45	30,93	30,62

**Cuadro C.1 – Mediciones de degradación para 12 escenas  
de la base de datos:  $\bar{m}_{i,j}^{(420)}$ ,  $\bar{m}_{i,j}^{(SIF)}$ ,  $\bar{F}_{i,j}$ ,  $\bar{G}_{i,j}$**

Escena j	$\bar{G}_{1,j}$ YP	$\bar{G}_{2,j}$ CbP	$\bar{G}_{3,j}$ CrP	$\bar{G}_{4,j}$ YE	$\bar{G}_{5,j}$ CbE	$\bar{G}_{6,j}$ CrE	$\bar{G}_{7,j}$ YT	$\bar{G}_{8,j}$ CbT	$\bar{G}_{9,j}$ CrT
1	1,85	4,17	3,80	1,27	4,38	4,61	1,63	5,36	4,99
2	3,52	8,17	10,04	1,50	8,26	8,80	2,36	8,63	10,54
3	3,69	6,52	9,67	2,09	6,73	7,65	2,35	8,25	9,32
4	2,84	10,70	8,95	1,15	5,56	5,24	1,31	5,23	5,06
5	5,25	14,15	12,98	2,00	10,02	8,41	3,37	13,05	12,69
6	4,07	19,74	11,09	1,26	10,68	7,10	1,63	11,86	7,54
7	4,42	8,98	8,96	1,81	9,17	6,63	2,08	7,32	6,29
8	2,19	8,71	10,86	1,69	8,74	8,73	2,64	8,97	9,20
9	3,11	6,45	7,26	2,33	5,69	8,05	2,69	5,66	5,83
10	6,49	15,92	10,79	2,02	9,03	8,06	2,57	7,88	7,96
11	5,50	4,71	5,78	1,50	2,27	3,80	1,78	3,70	3,86
12	13,04	9,30	10,01	44,95	40,64	32,98	45,45	30,93	30,62

### C.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 C-1 y C-2,  $i = 1, 2, \dots, 9$ . En este apartado se describe el cálculo de dichos parámetros tomando como referencia las escenas degradadas  $I^{(420)}$  e  $I^{(SIF)}$ .

#### C.7.1 Cálculo de $W_i$

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

$$D_{i,j} = \frac{1}{2} \cdot \left( \left| L_{i,j}^{(420)} - \bar{L}_{i,j}^{(420)} \right| + \left| L_{i,j}^{(SIF)} - \bar{L}_{i,j}^{(SIF)} \right| \right) \quad (C-8)$$

donde:

$$\begin{cases} \bar{L}_{i,j}^{(420)} = 100 / \left[ 1 + \left( \bar{F}_{i,j} / \bar{m}_i^{(420)} \right)^{\bar{G}_{i,j}} \right] \\ \bar{L}_{i,j}^{(SIF)} = 100 / \left[ 1 + \left( \bar{F}_{i,j} / \bar{m}_i^{(SIF)} \right)^{\bar{G}_{i,j}} \right] \\ L_{i,j}^{(420)} = 100 / \left[ 1 + \left( \bar{F}_{i,j} / m_i^{(420)} \right)^{\bar{G}_{i,j}} \right] \\ L_{i,j}^{(SIF)} = 100 / \left[ 1 + \left( \bar{F}_{i,j} / m_i^{(SIF)} \right)^{\bar{G}_{i,j}} \right] \end{cases} \quad (C-9)$$

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

$$a_k = \begin{cases} 1, & \text{si (escena k) } \in \Omega, \\ 0, & \text{en los demás casos} \end{cases} \quad (C-10)$$

$$W_{i,j} = \frac{a_j \cdot D_{i,j}^{-1}}{\sum_{k=1}^{12} a_k \cdot D_{i,k}^{-1}} \quad (\text{C-11})$$

Sea ahora  $i = \{1, 2, \dots, 9\} \equiv \{(plana, Y), (plana, Cb), (plana, Cr), (bordes, Y), (bordes, Cb), (bordes, Cr), (textura, Y), (textura, Cb), (textura, Cr)\}$ , siendo  $(plana, C)$ ,  $(bordes, C)$  y  $(textura, C)$  las regiones planas, de bordes y de textura de la componente de imagen  $C$ ,  $C = Y, Cb, Cr$ .

Sea  $u = textura, bordes, plana$  y  $v = Y, Cb, Cr$ , se calcularán los valores  $W_i$ ,  $i = 1, 2, \dots, 9$ , como:

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

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

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

$$W_i = \frac{\kappa_i}{\tau} \cdot \frac{1}{E_i} \quad (\text{C-12})$$

### C.7.2 Cálculo de $F_i$ y $G_i$

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

$$L_i^{(420)} = \frac{1}{\gamma} \cdot \sum_{j=1}^{12} W_{i,j} \cdot L_{i,j}^{(420)} \quad (\text{C-13})$$

$$L_i^{(SIF)} = \frac{1}{\gamma} \cdot \sum_{j=1}^{12} W_{i,j} \cdot L_{i,j}^{(SIF)} \quad (\text{C-14})$$

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  $(S_j, T_j)$  y  $(\bar{S}_j, \bar{T}_j)$ , de la escena de entrada y de cada escena de la base de datos, respectivamente. Los atributos espaciales  $\bar{S}_j$  de la base de datos de modelos de degradación se calculan directamente (véase la cláusula C.5) a partir de  $\bar{m}_{i,j}^{(420)}$ ,  $\bar{m}_{i,j}^{(SIF)}$ ,  $i = 1, 2, \dots, 9$  y  $j = 1, 2, \dots, 12$ .

$$D_j = (S - \bar{S}_j)^2 + (T - \bar{T}_j)^2 \quad (\text{C-15})$$

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

$$a = \sum_{j=1}^{12} w_j \cdot \left[ \frac{\bar{S}_j \cdot \bar{T}_j}{2} + (1 - \bar{T}_j^2) \cdot \left( 1 - \frac{\bar{S}_j^2}{2} \right) \right] \quad (C-16)$$

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

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

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

$$L_i^{(420)} = 100 / \left[ 1 + \left( \frac{F_i}{m_i^{(420)}} \right)^{G_i} \right] \quad (C-17)$$

$$L_i^{(SIF)} = 100 / \left[ 1 + \left( \frac{F_i}{m_i^{(SIF)}} \right)^{G_i} \right] \quad (C-18)$$

## C.8 Referencias

- [C-1] Recomendación UIT-R BT.500-11 (2000), *Metodología de evaluación subjetiva de la calidad de las imágenes de televisión*.
- [C-2] Recomendación UIT-R BT.802-1 (1994), *Imágenes y secuencias de prueba para las evaluaciones subjetivas de códecs digitales que cursan señales producidas conforme a la Recomendación UIT-R BT.601*.
- [C-3] Material de referencia UIT-T (2004), *Objective perceptual assessment of video quality: Full reference television*.
- [C-4] Recomendación UIT-R BT.601 (1995), *Parámetros de codificación de televisión digital para estudio con formatos de imagen normal 4:3 y de pantalla ancha 16:9*.
- [C-5] Recomendación UIT-T H.262 (2000), *Tecnología de la información – Codificación genérica de imágenes en movimiento e información de audio asociada: Video*.
- [C-6] ISO/IEC 11172-1:1993, *Information technology – Coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbit/s – Part 1: Systems*.
- [C-7] Gonzalez, R.C. and Woods, R.E. (1992), *Digital Image Processing*, Addison-Wesley.
- [C-8] Trucco, E. and Verri A. (1998), *Introductory Techniques for 3-D Computer Vision*, Prentice-Hall.

## C.9 Resultados objetivos de las pruebas VQEG-fase II

### Cuadro C.2 – Datos objetivos brutos 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				0,7036		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 C.3 – Datos objetivos brutos 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*	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

\* Se ha suprimido del análisis el valor SRC = 6, HRC = 5 por superar los requisitos de registro temporal del plan de la prueba VQEG.

## Anexo D

### **Administración nacional de telecomunicaciones e información (NTIA, *national telecommunications and information administration*)**

#### **Descripción técnica de la métrica de la calidad de vídeo (VQM, *video quality metric*)**

En este anexo se presenta la descripción funcional completa de la VQM de la NTIA y de las técnicas de calibración asociadas. Los algoritmos de calibración descritos en este anexo bastan para asegurar 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  $\frac{1}{2}$  píxel y una precisión de registro temporal de más menos un campo entrelazado.

#### **D.1 Introducción**

En este anexo se presenta una descripción técnica completa del modelo general de la Administración nacional de telecomunicaciones e información (NTIA, *national telecommunications and information administration*) 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 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  $\frac{1}{2}$  píxel y una precisión de registro temporal de más menos un campo entrelazado.

La NTIA se ha mostrado dispuesta a suministrar a todas las partes interesadas una aplicación informática que aplica el Modelo general y las técnicas de calibración asociadas. Las partes interesadas podrán descargarla de: [www.its.blrdoc.gov/n3/video/vqmsoftware.htm](http://www.its.blrdoc.gov/n3/video/vqmsoftware.htm).

Descargo de responsabilidad: En ningún caso será la UIT responsable por los daños de cualquier naturaleza (incluidos, sin limitación, los daños causados por lucro cesante, interrupción de la actividad comercial, pérdida de información u otras pérdidas pecuniarias) resultantes o relacionadas con el uso o imposibilidad de uso del programa en cuestión. La UIT rehúsa cualquier garantía, explícita o implícita, incluidas, pero no exclusivamente, las garantías de comercialidad e idoneidad para un fin determinado.

#### **D.2 Referencias**

##### **D.2.1 Referencias normativas**

- Recomendación UIT-R BT.601-5 (1995), *Parámetros de codificación de televisión digital para estudios con formatos de imagen normal 4:3 y de pantalla ancha 16:9.*

### D.3 Definiciones

**D.3.1 4:2:2:** Formato de muestreo de imágenes Y, Cb, Cr en el que los planos de crominancia (Cb y Cr) se muestrean horizontalmente a la mitad de velocidad de muestreo del plano de luminancia (Y). Véase la Rec. UIT-R BT.601-5 (cláusula D.2).

**D.3.2 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 D.7.5 su definición matemática exacta.

**D.3.3 YUV grande:** Formato de fichero binario utilizando para almacenar cortos que se hayan muestreado con arreglo a la Rec. UIT-R BT.601-5. 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 Rec. UIT-R BT.601-5. Y representa la información del canal de luminancia, U representa el canal de diferencia de color azul (es decir  $C_B$  en la Rec. UIT-R BT.601-5), y V representa el canal de diferencia de color rojo (o sea  $C_R$  en la Rec. UIT-R BT.601-5). El orden de los píxeles en el fichero binario coincide con el especificado en SMPTE 125M [D-7]. La especificación completa del formato de fichero YUV grande Figura en la cláusula D.5 mientras que las rutinas de software para la lectura y visualización de los ficheros YUV grandes Figura en [D-14].

**D.3.4 corto:** Representación digital de una escena almacenada en un medio informático.

**D.3.5 VQM de corto:** VQM de un solo corto de vídeo procesado.

**D.3.6 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 ( $C_R$ ).

**D.3.7 códec:** Abreviatura de codificador/decodificador o de compresor/descompresor.

**D.3.8 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 [D-2].

**D.3.9 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.

**D.3.10 campo:** Mitad de un cuadro, conteniendo todas las líneas pares o impares.

**D.3.11 cuadro:** Una imagen completa de televisión.

**D.3.12 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.

**D.3.13 ganancia:** Factor de escala aplicado por el circuito ficticio de referencia (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.

**D.3.14 modelo general:** Modelo de calidad de vídeo, o VQM, objeto de este anexo cláusula D.9. El modelo general se sometió a las pruebas de la fase 2 llevadas a cabo por el Grupo de expertos en calidad de vídeo (VQEG). El Informe final del VQEG fase-2 describe la calidad de funcionamiento del modelo general (véase [D-15], proponente H).

**D.3.15 H.261:** Abreviatura de la Recomendación UIT-T H.261 [D-2].

**D.3.16 circuito ficticio de referencia (HRC, *hypothetical reference circuit*):** Sistema de vídeo objeto de la prueba tal como un códec o un sistema digital de transmisión de vídeo.

**D.3.17 vídeo de entrada:** Vídeo antes de ser procesado o distorsionado por un HRC (véase la Figura D.1). El vídeo de entrada puede denominarse también vídeo original.

**D.3.18 unidad Instituto de ingenieros de radiocomunicaciones (IRE, *Institute for radio engineers*):** Unidad de tensión que se utiliza comúnmente para la medición de señales de vídeo. Un IRE es equivalente a 1/140 de voltio.

**D.3.19 Unión Internacional de Telecomunicaciones (UIT):** Organización Internacional del sistema de Naciones Unidas en la que los Estados y el sector privado coordinan las redes y servicios de telecomunicaciones mundiales. La UIT comprende: Unión Internacional de Telecomunicaciones – Sector de Radiocomunicaciones (UIT-R) y Unión Internacional de Telecomunicaciones – Sector de Normalización de las Telecomunicaciones (UIT-T).

**D.3.20 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).

**D.3.21 nota media de opinión (MOS, *mean opinion score*):** Juicio medio de calidad subjetiva otorgado por un grupo especial de espectadores a un corto de vídeo procesado.

**D.3.22 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).

**D.3.23 Comisión nacional de sistemas de televisión (NTSC, *national television systems committee*):** Sistema compuesto de vídeo analógico en color de 525 líneas [D-8].

**D.3.24 desplazamiento o desplazamiento de nivel:** Sumando aplicado por el circuito ficticio de referencia (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.

**D.3.25 región original de interés (OROI, *original region of interest*):** Región de interés (ROI) extraída del vídeo original, definida por sus coordenadas rectangulares.

**D.3.26 vídeo original:** Vídeo antes de ser procesado o distorsionado por el HRC (véase la Figura D.1). El vídeo original recibe además el nombre de vídeo de entrada por ser la entrada de vídeo al sistema digital de transmisión de vídeo.

**D.3.27 región válida original (OVR, *original valid region*):** Región válida de un corto de vídeo original, definida por sus coordenadas rectangulares.

**D.3.28 vídeo de salida:** Vídeo procesado o distorsionado por el HRC (véase la Figura D.1). El vídeo de salida recibe también el nombre de vídeo procesado.

**D.3.29 sobreexploración:** Porción de vídeo que no suele ser visible en una pantalla de televisión normal.

**D.3.30 líneas con alternancia de fase (PAL, *phase alternation lines*):** Sistema compuesto de vídeo analógico en color de 625 líneas.

**D.3.31 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.

**D.3.32 región de interés procesada (PROI, *processed region of interest*):** Región de interés (ROI) extraída del vídeo procesado y con los desplazamientos espaciales del HRC corregidos, definida por sus coordenadas rectangulares.

**D.3.33 vídeo procesado:** Vídeo procesado o distorsionado por el HRC (véase la Figura D.1). El vídeo procesado recibe también el nombre de vídeo de salida por ser la salida de vídeo del sistema digital de transmisión de vídeo.

**D.3.34 región válida procesada (PVR, *processed valid region*):** Región válida de un corto de vídeo procesado del 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 región válida original (OVR). La región entre la PVR y la OVR es la porción de vídeo borrada o corrompida por el HRC.

**D.3.35 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 Rec. UIT-R BT.601-5 la abertura 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 [D-9].

**D.3.36 cuarto de 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 [D-2].

**D.3.37 Rec. UIT-R BT.601-5:** Abreviatura de la Rec. UIT-R BT.601-5 (cláusula D.2), norma 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 la cláusula D.5 para más información.

**D.3.38 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 D.5.3.

**D.3.39 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 [D-3]. 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.

**D.3.40 reencuadre:** 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 D.6.1.2.

**D.3.41 región de interés (ROI, *region of interest*):** Retícula de imagen (definida por sus coordenadas rectangulares) utilizada para indicar una subregión particular de un campo o cuadro de vídeo. Véase asimismo SROI.

**D.3.42 escena:** Secuencia de cuadros de vídeo.

**D.3.43 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 [D-4] contiene una definición de SI basada en estadísticas extraídas de  $3 \times 3$  imágenes filtradas con Sobel [D-6] mientras que D.7.2.2 presenta una definición de SI basada en estadísticas extraídas de  $13 \times 13$  imágenes mucho más grandes filtradas en los bordes (Figura D.11).

**D.3.44 región espacial de interés (SROI, *spatial region of interest*):** Retícula 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 Rec. UIT-R BT.601-5, 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 ROI.

**D.3.45 registro espacial:** Proceso utilizado para estimar y corregir los desplazamientos espaciales de la secuencia de vídeo procesada con respecto a la secuencia de vídeo original.

**D.3.46 subregión espacial-temporal (S-T, *spatial-temporal 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 Figura D.9.

**D.3.47 Sociedad de ingenieros de imágenes en movimiento y televisión (SMPTE, *society of motion picture and television engineers*):** Principal 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.

**D.3.48 información temporal (TI, *temporal information*):** Característica basada en estadísticas extraídas de los gradientes temporales (es decir del movimiento) de una escena de vídeo. La referencia [D-4] y D.7.5 contienen definiciones de la TI basada en estadísticas extraídas de diferencias entre cuadros sencillos.

**D.3.49 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.

**D.3.50 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 D.6.4.1).

**D.3.51 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.

**D.3.52 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 D.6.4.

**D.3.53 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.

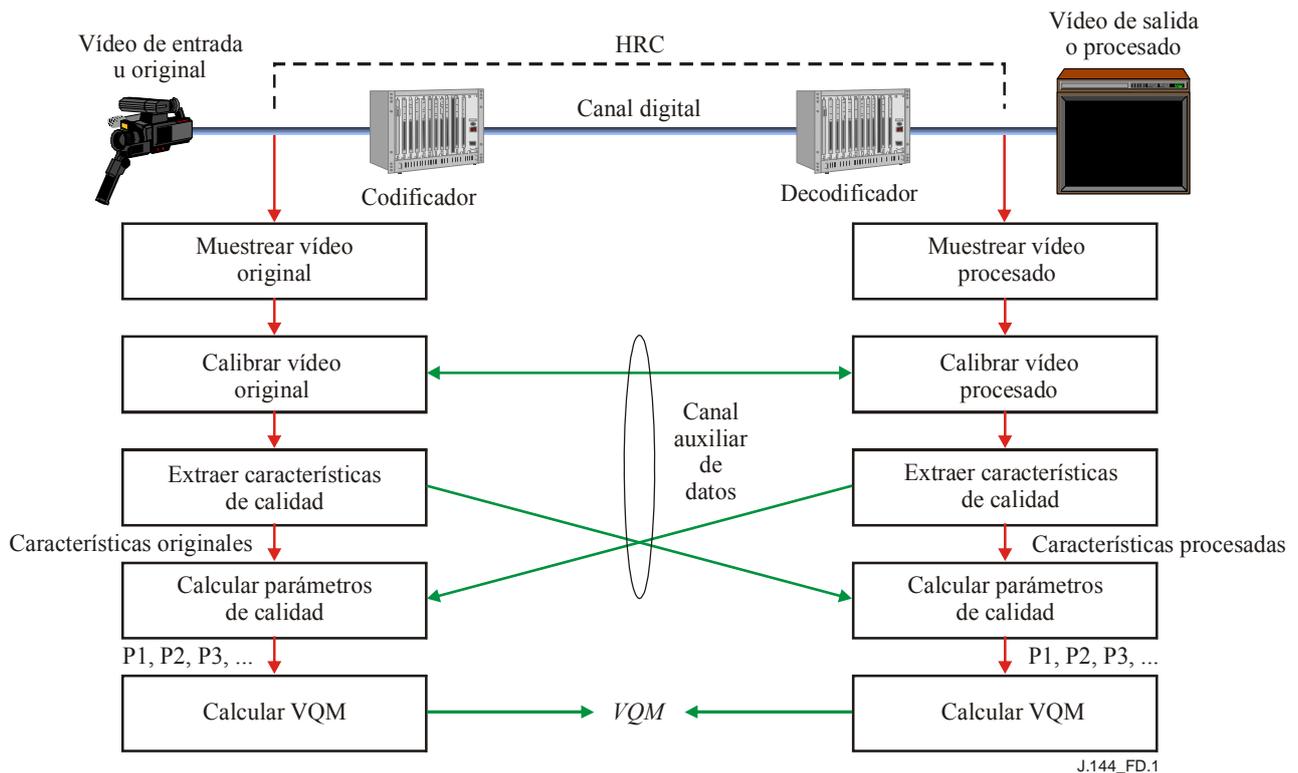
**D.3.54 grupo de expertos en calidad de vídeo (VQEG, *video quality experts group*):** Grupo de expertos internacionales en calidad de vídeo que realizan las pruebas de validación de las métricas objetivas de calidad de funcionamiento de vídeo. Los resultados del VQEG se comunican a la Unión Internacional de Telecomunicaciones (UIT) y pueden utilizarse como base para las Recomendaciones internacionales de medición de la calidad de vídeo.

**D.3.55 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.

## D.4 Esquema del cálculo de la métrica de calidad del vídeo (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 de opinión medias, de las degradaciones de vídeo digital sometidas a la calificación de grupos especiales de espectadores [D-1]. La Figura D.1 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 (cláusula D.5), la calibración de las secuencias de vídeo original y procesada (cláusula D.6), la extracción de las características basadas en la percepción (cláusula D.7), el cálculo de los parámetros de calidad del vídeo (cláusula D.8), y el cálculo del modelo general (cláusula D.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).

El método de medición presentado en este documento utiliza parámetros de referencia reducida de anchura de banda alta [D-3]. Estos parámetros de referencia reducida utilizan características extraídas de las regiones espacio-temporales (S-T) de la secuencia de vídeo (véase D.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 Figura D.1.



**Figura D.1 – Pasos necesarios para calcular la VQM**

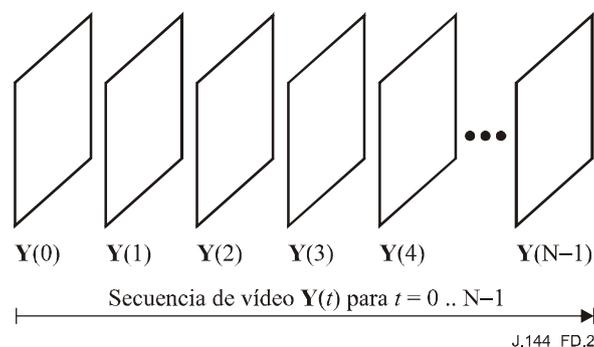
## D.5 Muestreo

Los algoritmos de este anexo se han diseñado para ser ejecutados en un computador en la hipótesis de que las secuencias de vídeo original y procesada están 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 Rec. UIT-R BT.601-5 (cláusula D.2). El vídeo compuesto, tal como el NTSC o el PAL, debe

convertirse antes en vídeo de componentes con las tres señales siguientes: luminancia ( $Y$ ), diferencia de color azul ( $C_B$ ), y diferencia de color rojo ( $C_R$ ). El muestreo de la Rec. UIT-R BT.601-5 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 Rec. UIT-R BT.601-5 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 Rec. UIT-R BT.601-5 tendrá 720 píxeles por 486 líneas. Análogamente, cuando se muestrea vídeo PAL de 625 líneas con arreglo a la Rec. UIT-R BT.601-5, el cuadro de vídeo  $Y$  contendrá 720 píxeles por 576 líneas. Si se utilizan 8 bits para muestrear uniformemente la señal  $Y$ , la Rec. UIT-R BT.601-5 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".

#### D.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 Rec. UIT-R BT.601-5, 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 Figura D.1,  $t = 0$  indica el primer cuadro muestreado mientras que  $t = (N - 1)$  indica el último.



**Figura D.2 – Indexación temporal de los cuadros en los ficheros YUV grandes**

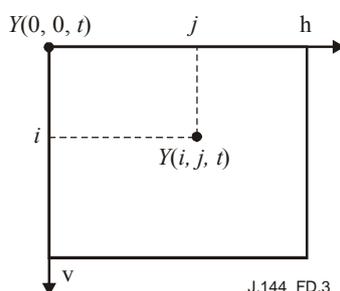
Todos los algoritmos se desarrollan y describen para su ejecución con pares de ficheros muestreados: una secuencia de vídeo original y su secuencia asociada de vídeo procesado. Para evitar confusiones, se supone que ambos ficheros tienen la misma longitud. Además, se establece la hipótesis inicial de 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 Figura D.1. Por consiguiente, la incertidumbre restante de la estimación de retardo de vídeo se representará por más menos  $U$ .

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

En la Figura D.3 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  $Y(i, j, t)$ .

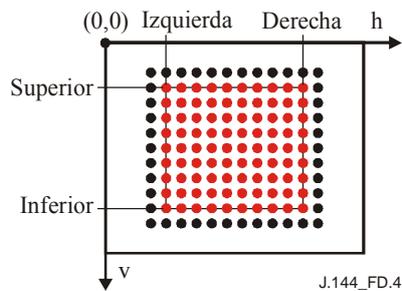


**Figura D.3 – Sistema de coordenadas utilizado en los cuadros de luminancia muestreada Y**

Un corto de vídeo muestreado con arreglo a la Rec. UIT-R BT.601-5 se guarda en el formato de fichero "YUV grande", donde  $Y$  representa la información de luminancia de la Rec. UIT-R BT.601-5,  $U$  representa la información de diferencia de color azul (es decir  $C_B$  en la Rec. UIT-R BT.601-5), y  $V$  representa la información de diferencia de color rojo (es decir  $C_R$  en la Rec. UIT-R BT.601-5). 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 [D-7].

### D.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 Figura D.4 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 Figura D.4 permite definir la rejilla utilizada para mostrar estas subregiones contiguas y los cálculos realizados para extraer las características de cada subregión contigua.



**Figura D.4 – Coordenadas rectangulares para la especificación de subregiones de imagen**

#### **D.5.4 Consideraciones sobre las secuencias de vídeo de más de 10 segundos**

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 segundos. 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.

#### **D.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:

- 1) estimación y corrección del registro espacial;
- 2) estimación de la región válida para limitar la extracción de características a aquellos píxeles que contengan información de la imagen;
- 3) estimación y corrección del desplazamiento de nivel y ganancia (normalmente conocidos como contraste y brillo); y
- 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 [D-5] 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 (cláusulas D.7 y D.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 D.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.

## **D.6.1 Registro espacial**

### **D.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 D.6.1.6.

El registro espacial debe determinarse antes que la región válida procesada (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 tanto 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.

#### **D.6.1.2 Cuestiones de entrelazado**

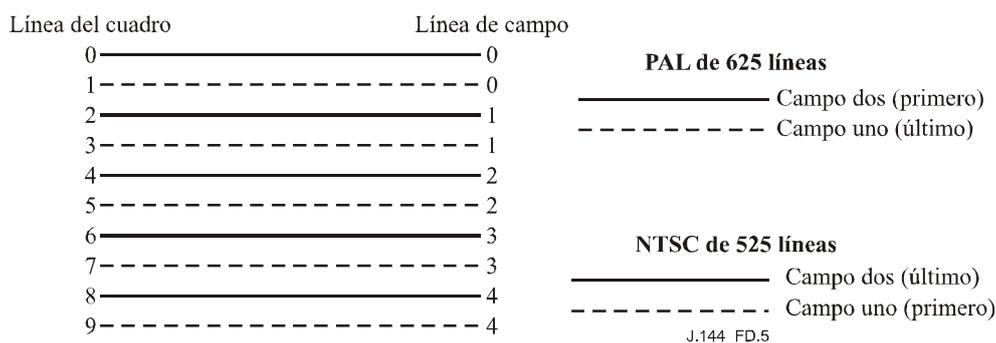
El registro espacial vertical del vídeo entrelazado constituye un problema mayor que el del vídeo progresivo, ya que el proceso de registro espacial debe distinguir entre el campo uno y el campo dos. Hay tres condiciones de desplazamiento vertical que deben distinguirse para obtener un registro correcto 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 Figura D.5, 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.

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 de 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.



**Figura D.5 – Diagrama que muestra los campos entrelazados y el esquema de numeración de las líneas del cuadro/campo**

Si el desplazamiento vertical del campo dos no es igual a 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.

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

Esta cláusula 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.

#### **D.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 Rec. UIT-R BT.601-5 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 Rec. UIT-R BT.601-5 (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.

#### **D.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).

### **D.6.1.3.3 Hipótesis de la región válida procesada (PVR)**

La hipótesis de la región válida procesada (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 Rec. UIT-R BT.601-5, 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 Rec. UIT-R BT.601-5, 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.

### **D.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.

#### **D.6.1.4.1 Región de interés (ROI) utilizada en todos los cálculos**

El algoritmo efectuará todas las comparaciones de campos entre versiones de un ROI extraído 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 región válida procesada (PVR) para todos los posibles desplazamientos espaciales que se examinen.
- La OROI quedará centrada en la imagen original.

#### **D.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 D.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 D.6.1.4.3).

#### **D.6.1.4.3 Fórmulas utilizadas para comparar la PROI con la OROI**

Tras corregir la ganancia<sup>6</sup> de la PROI (véase D.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).

#### **D.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 arbitrario el algoritmo puede encontrar un registro espacial incorrecto (véase D.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). Esta cláusula 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.

##### **D.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.

---

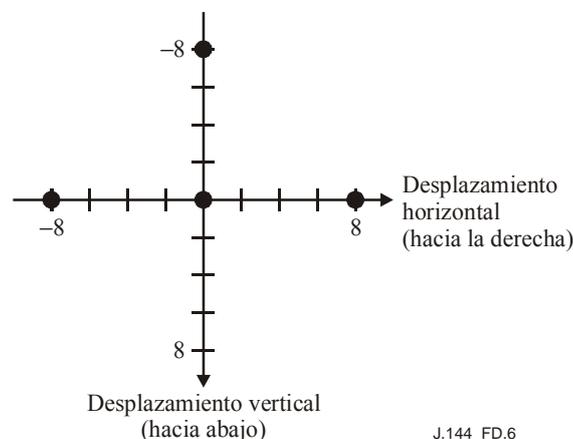
<sup>6</sup> A veces es posible omitir la compensación de ganancia para reducir la complejidad computacional. No obstante, sólo se recomienda la omisión de la corrección de ganancia durante las primeras etapas del algoritmo de búsqueda iterativo, cuando el objetivo es encontrar el registro espacial aproximado (véanse por ejemplo, D.6.1.5.2 y D.6.1.5.3).

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.

#### D.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.

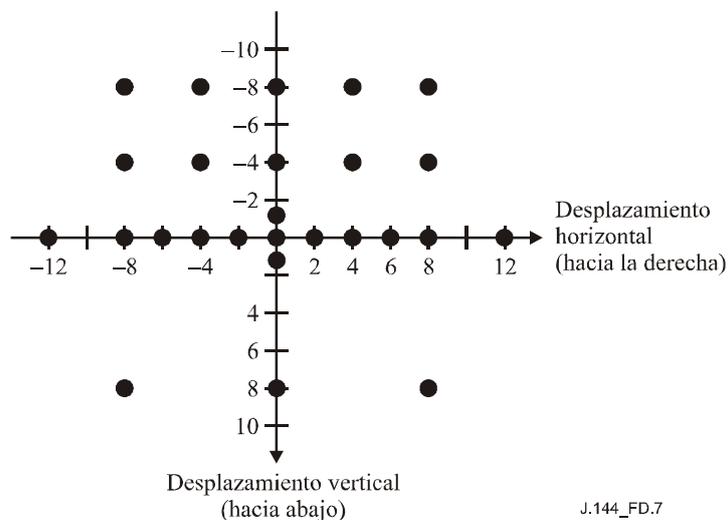
En esta búsqueda general se examina el campo uno del cuadro procesado seleccionado (véase la Figura D.5), considerándose únicamente los campos originales de tipo uno separados dos cuadros (es decir separados cuatro campos) en todo el intervalo de más menos la incertidumbre del registro temporal. La búsqueda 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 Figura D.6). En la Figura D.6, 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 D.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 D.6.1.5.3). De acuerdo con el sistema de coordenadas de la Figura D.3, 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 D.6 – Desplazamientos espaciales considerados en la búsqueda general del desplazamiento temporal**

### D.6.1.5.3 Búsqueda general del desplazamiento espacial

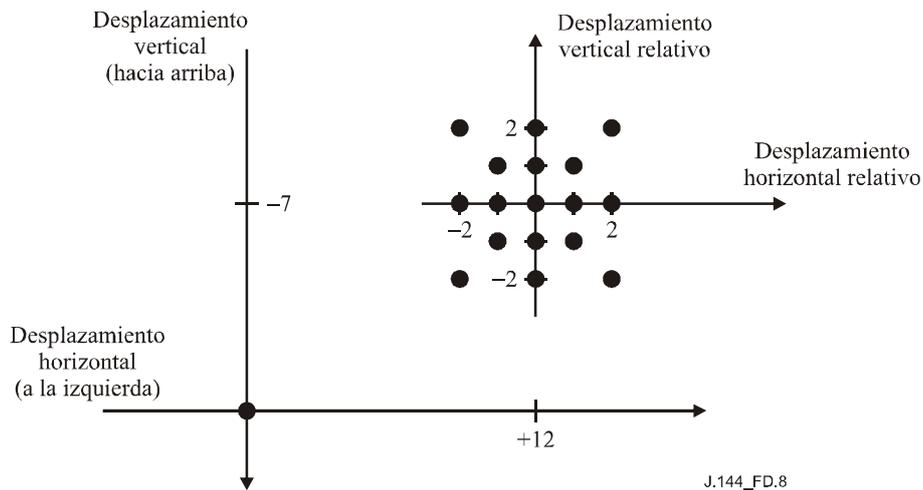
Con el registro temporal encontrado en la búsqueda general del desplazamiento temporal (véase D.6.1.5.2), se efectúa ahora una búsqueda general del desplazamiento espacial correcto con un rango más limitado de campos originales. El rango de campos originales considerado en esta búsqueda incluye el campo original tipo uno que más se ajusta (del D.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 Figura D.7. Obsérvese que se consideran menos desplazamientos hacia abajo (como en D.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 D.6.1.4.3. Los mejores desplazamientos temporal y espacial resultantes pasan a ser las estimaciones mejoradas para el siguiente paso del algoritmo del D.6.1.5.4.



**Figura D.7 – Desplazamientos espaciales considerados en la búsqueda general del desplazamiento espacial**

### D.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 Figura D.8). En el ejemplo de la Figura D.8, la estimación actual del desplazamiento espacial para el vídeo procesado es de siete líneas de campo hacia arriba y doce píxeles a la derecha del vídeo original. El conjunto de desplazamientos espaciales representado en la Figura D.8 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 D.6.1.4.3. Los mejores desplazamientos espaciales y temporales resultantes pasan a ser las estimaciones mejoradas para el siguiente paso del algoritmo del D.6.1.5.5.



**Figura D.8 – Desplazamientos espaciales considerados en la búsqueda detallada del desplazamiento espacial**

#### D.6.1.5.5 Búsquedas detalladas repetitivas

La iteración de la búsqueda detallada del D.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 compensación de ganancia, la ganancia del campo procesado se estimará de nuevo entre cada dos búsquedas detalladas (véase D.6.1.4.2).

Las búsquedas detalladas repetitivas se efectúan sobre el campo procesado (véase D.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 D.6.1.5.6) y para cada HRC se examinan varias escenas (véase D.6.1.5.7).

#### D.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 D.6.1.5.2). Obsérvese que esta búsqueda general se efectúa sobre los primeros  $U \times 2 + 1$  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 D.6.1.5.3). A continuación se efectúan hasta cinco búsquedas detalladas espacio-temporales para afinar las estimaciones espaciales y temporales (véanse D.6.1.5.4 y D.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>8</sup>.

#### **D.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.

#### **D.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.

#### **D.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 D.6.1.5.3) los números del eje vertical de la Figura D.7 deben duplicarse (o sea, +8 pasa a ser +16 y -4 pasa a ser -8)<sup>9</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 Figura D.7, 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 véase D.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

---

<sup>8</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.

<sup>9</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).

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 D.6.1.5.4. En ésta, los desplazamientos verticales quedan inalterados, y oscilan entre  $-2$  líneas de cuadro y  $+2$  líneas de cuadro. Por consiguiente, el eje vertical de la Figura D.8 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.

## **D.6.2 Región válida**

El vídeo NTSC (de 525 líneas) y el PAL (de 625 líneas) muestreados con arreglo a la Rec. 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 Rec. UIT-R BT.601-5. 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 Rec. UIT-R BT.601-5, los píxeles de los bordes negros ni los efectos de transición cuando los bordes negros alcanzan la zona de imagen.

### D.6.2.1 Algoritmo central de la región válida

En esta cláusula 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 Rec. UIT-R BT.601-5 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 registro espacial en D.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 D.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.

- 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 " $M_{J-1}$ ".
- 2) Se toma la media de la siguiente columna de píxeles, " $M_J$ ".
- 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).
- 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.

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

### **D.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 Rec. UIT-R BT.601-5 (cláusula D.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 Rec. UIT-R BT.601-5, 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 Rec. UIT-R BT.601-5, 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.

### **D.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.

### **D.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 Rec. UIT-R BT.601-5 (es decir de formato intermedio común, o CIF, utilizado por la Rec. 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.

## **D.6.3 Ganancia y desplazamiento**

### **D.6.3.1 Algoritmo central de ganancia y desplazamiento de nivel**

Esta cláusula 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 D.6.1). Las imágenes original y procesada deben estar asimismo registradas temporalmente, lo que se abordará más adelante en D.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 Rec. UIT-R BT.601-5 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 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:

$$\underline{Q}_{N \times 1} = \begin{bmatrix} original_1 \\ \cdot \\ \cdot \\ \cdot \\ original_N \end{bmatrix}, \quad \underline{P}_{N \times 1} = \begin{bmatrix} procesado_1 \\ \cdot \\ \cdot \\ \cdot \\ procesado_N \end{bmatrix}$$

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

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

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  $A_{N \times 2} = [\underline{1} \quad \underline{Q}]$ , 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  $C_R$ .

### D.6.3.2 Utilización de escenas

El algoritmo básico del D.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 región válida procesada (PVR).

### D.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 D.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 D.6.3.2, tras corregir el desplazamiento espacial del vídeo procesado. Utilizando las imágenes  $Y$  y submuestreadas, se aplicara a la función de búsqueda del D.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  $C_R$ .

### D.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 [ $Y$ ,  $C_B$ ,  $C_R$ ].

- 1) Utilizar la solución normal de mínimos cuadrados del D.6.3.1 para generar la estimación inicial del desplazamiento de nivel y ganancia: 
$$\begin{bmatrix} l \\ g \end{bmatrix} = (A^T A)^{-1} A^T \underline{P}.$$
- 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:  $\underline{E} = |\underline{P} - \hat{\underline{P}}|.$
- 3) Generar un vector de costos ( $\underline{C}$ ) que sea recíproco, elemento a elemento, del vector de errores ( $\underline{E}$ ) más un pequeño épsilon ( $\epsilon$ ):  $\underline{C} = \frac{1}{\underline{E} + \epsilon}$ . 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.
- 4) Normalizar el vector de costos  $\underline{C}$  para la norma unitaria (es decir cada elemento de  $\underline{C}$  se divide por la raíz cuadrada de la suma de los cuadrados de todos los elementos de  $\underline{C}$ ).
- 5) Generar el vector de costos  $\underline{C}^2$  que sea el cuadrado, elemento a elemento, del vector de costos  $\underline{C}$  del paso 4.
- 6) Generar una matriz diagonal  $N \times N$  de costos ( $C^2$ ) que contenga los elementos del vector de costos ( $\underline{C}^2$ ) dispuesto sobre la diagonal, relleno el resto a ceros.
- 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: 
$$\begin{bmatrix} l \\ g \end{bmatrix} = (A^T C^2 A)^{-1} A^T C^2 \underline{P}.$$
- 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.

#### **D.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.

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

Los algoritmos de registro temporal (véase D.6.4) y la mayor parte de las características de calidad (cláusula D.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:

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

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.

#### **D.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 Figura D.1 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. Esta cláusula 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.

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

En esta cláusula 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.

##### **D.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.

##### **D.6.4.1.2 Entradas al algoritmo**

Una secuencia de N imágenes de luminancia de vídeo original:  $\mathbf{Y}_O(t)$ ,  $0 \leq t < N^{10}$ .

Una secuencia de N imágenes de luminancia de vídeo procesado:  $\mathbf{Y}_P(t)$ ,  $0 \leq t < N$ .

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).

---

<sup>10</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 Figura D.2.

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  $0 \leq t < N$ .

#### D.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.

#### D.6.4.1.4 Descripción del algoritmo

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.

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.

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.

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>11</sup>. Esta normalización

---

<sup>11</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.

minimizará la influencia de las 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.

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  $\mathbf{C}_{td}$  y se calcularán como la desviación típica en el espacio de la imagen formada restando la imagen procesada  $t$  de la imagen original  $t+d$ :  $\mathbf{C}_{td} = std_{space}(\mathbf{S}_O(t+d) - \mathbf{S}_P(t))$ . Estas comparaciones,  $\mathbf{C}_{td}$ , correlacionan la  $t$ -ésima imagen procesada con cada imagen original que esté dentro de la incertidumbre de registro. Valores inferiores de  $\mathbf{C}_{td}$  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.

6) *Efectuar una comprobación global de vídeo en reposo*

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

$$A_d = \frac{1}{N - 2 \times U} \sum_{t=U}^{N-U-1} \mathbf{C}_{td}$$

Esta suma 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.

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  $\mathbf{C}_{td}$ . Dicho de otro modo, para cada imagen procesada  $t$ , encontrar  $d_{\min}(t)$  tal que  $\mathbf{C}_{t \ d_{\min}(t)} \leq \mathbf{C}_{td}$ , 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.).

8) *Ejecutar una comprobación de reposo en cada imagen procesada*

Si para una determinada imagen  $t$  y todos los valores de  $d$  ( $-U \leq d \leq U$ ),  $\text{máximo}(\mathbf{C}_{td}) - \text{mínimo}(\mathbf{C}_{td}) < \text{STILL\_THRESHOLD}$ ,  $d_{\min}(t)$  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.

9) *Construir un histograma de todos los registros temporales definidos*

Calcular un histograma con los valores definidos de  $d_{\min}(t)$  con  $2 \times U + 1$  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  $+U$ .

10) *Formar un histograma suavizado*

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

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

Cuando se considera el histograma suavizado,  $SH_d$ , 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.

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 ( $H_d > \text{max\_H\_value} \times \text{BELOW\_WARN}$ ), 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 ( $-U \leq d < -U + \text{HFW}$ ) y ( $U - \text{HFW} < d \leq U$ ).
- ¿Tiene  $SH_d$  un retardo bien definido? Examinar  $SH_d$ , excepto dentro de DELTA de  $\text{max\_SH\_offset}$ . Si ( $SH_d > \text{max\_SH\_value} \times \text{BELOW\_WARN}$ ) 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.

#### D.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 (paso 11 del D.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.

#### **D.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.

### **D.7 Características de calidad**

#### **D.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* (cláusula D.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 prominencia 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 las cláusulas 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 la cláusula D.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>12</sup>.

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].

- 1) [Opcional] Aplicar un filtro perceptual.
- 2) Dividir la secuencia de vídeo en regiones S-T.

---

<sup>12</sup> Existe la hipótesis implícita de que la distancia de observación en 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 la cláusula D.9 otras observaciones relativas a la distancia de observación supuesta.

- 3) Extraer las características, o estadísticas resumen, de cada región S-T (por ejemplo, media y desviación típica).
- 4) [Opcional] aplicar un umbral de perceptibilidad.

Determinadas características pueden utilizar dos o más filtros perceptuales distintos.

#### **D.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:

- 1) el número de píxeles horizontalmente;
- 2) el número de líneas de cuadro verticalmente; y
- 3) la duración temporal de la región, expresada en unidades de cuadros de vídeo equivalentes con referencia a un sistema de vídeo de 30 cuadros por segundo (fps)<sup>13</sup>.

En la Figura D.9 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.

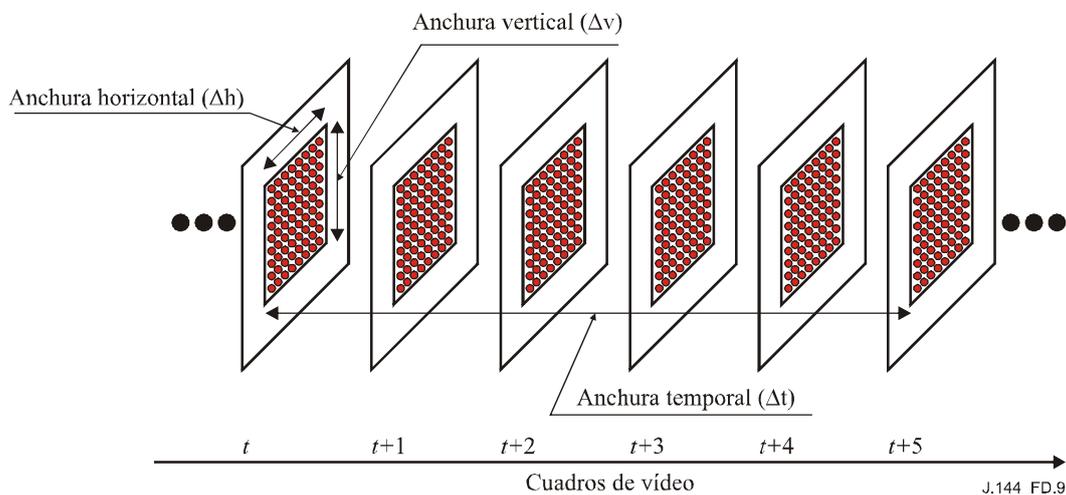
La región espacial de interés (SROI, véase la cláusula D.3) ocupada por todas las regiones S-T es idéntica para las secuencias de vídeo original y procesada calibrada. La SROI debe quedar totalmente dentro de la PVR, posiblemente con una zona intermedia de píxeles necesaria para cualquier filtro perceptual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

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.

---

<sup>13</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.

Para algunas características tales como las presentadas en D.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 [D-12].



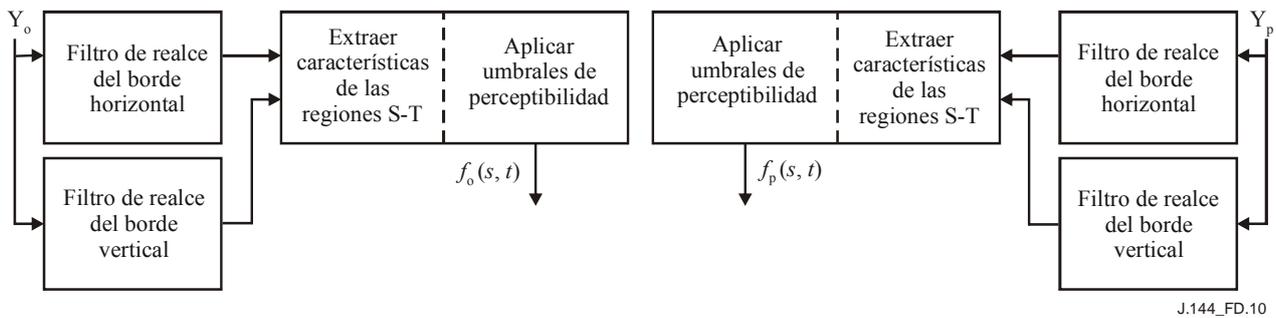
**Figura D.9 – Ejemplo de tamaño de región espacio-temporal (S-T) para extraer las características**

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.

### D.7.2 Características basadas en gradientes espaciales

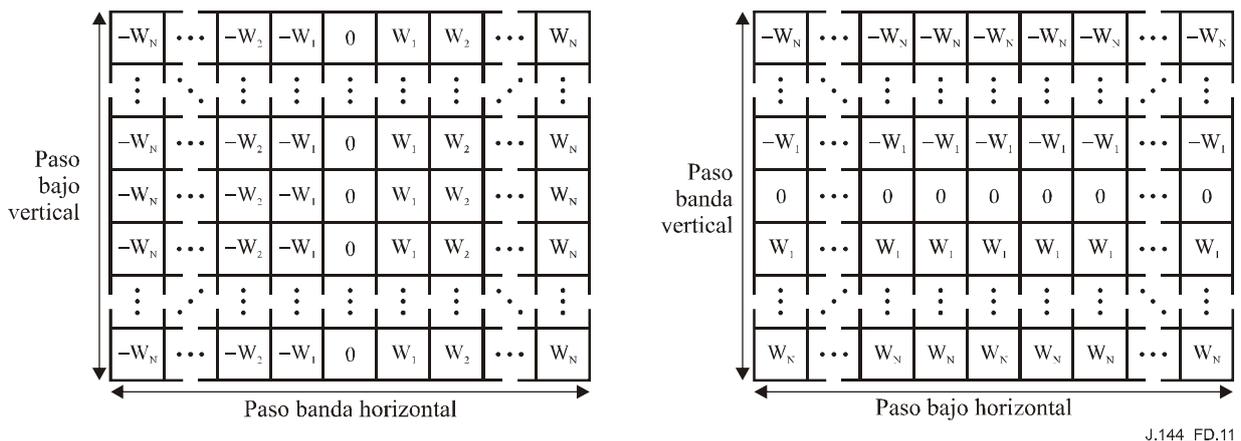
Se pueden utilizar características derivadas de gradientes espaciales para caracterizar las distorsiones perceptuales de los bordes. Por ejemplo, el difuminado provoca una pérdida general de información de los bordes mientras que un exceso de información de los bordes horizontales y verticales puede 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 espacio-temporales (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 Rec. UIT-R BT.601-5, evaluados subjetivamente a una distancia de observación de seis alturas de imagen. En la Figura D.10 se presenta un esquema del algoritmo utilizado para extraer las características basadas en gradientes espaciales.



**Figura D.10 – Esquema del algoritmo utilizado para extraer las características del gradiente espacial**

### D.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 Figura D.11 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 D.11 – Filtros de realce de bordes**

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

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

siendo  $x$  el desplazamiento en píxeles del centro del filtro (0, 1, 2, ..., N),  $c$  una constante que determina la anchura del filtro paso banda, y  $k$  una constante de normalización seleccionada de modo que cada filtro produzca la misma ganancia como un auténtico filtro Sobel [6]. 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 Figura D.11 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 Figura D.11, sólo es necesario sumar los píxeles de una columna y multiplicar una vez por el factor de ponderación).

### D.7.2.2 Descripción de las características $f_{SI13}$ y $f_{HV13}$

En esta cláusula 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 D.7.2.1. Se utilizarán estas características para detectar degradaciones espaciales tales como la borrosidad y el bloqueo. El filtro representado en la Figura D.11 (a la izquierda) realiza los gradientes espaciales en sentido horizontal (H) mientras que el transpuesto de este filtro (a la derecha) realiza 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 Figura D.12 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:

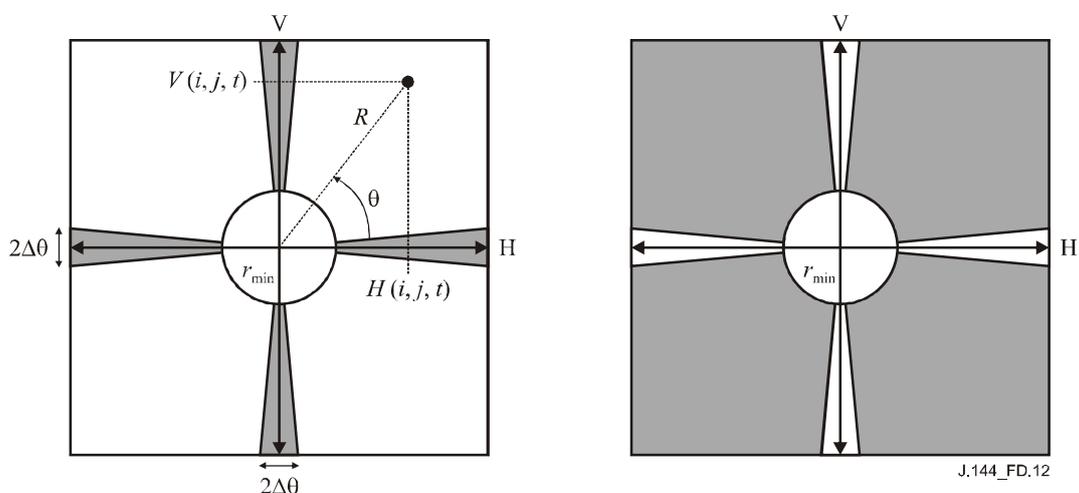
$$R(i, j, t) = \sqrt{H(i, j, t)^2 + V(i, j, t)^2}, \text{ y}$$

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

La primera característica es una medida de la información espacial (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 Figura D.11. 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:

$$f_{SI13} = \{std[R(i, j, t)]\}_P : i, j, t \in \{\text{Región S-T}\}$$

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.



**Figura D.12 – División de la actividad espacial horizontal (H) y vertical (V) en distribuciones  $HV$  (izquierda) y  $\overline{HV}$  (derecha)**

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 Figura D.12. 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 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:

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

y

$$\overline{HV}(i, j, t) = \left\{ \begin{array}{l} R(i, j, t) \text{ if } R(i, j, t) \geq r_{\min} \text{ y } m\frac{\pi}{2} + \Delta\theta \leq \theta(i, j, t) \leq (m+1)\frac{\pi}{2} - \Delta\theta \quad (m=0,1,2,3) \\ 0 \text{ en los demás casos} \end{array} \right\} \text{ donde}$$

donde:

$$i, j, t \in \{\text{Región S-T}\}$$

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:

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

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>14</sup>.

<sup>14</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).

### D.7.3 Características basadas en la información de crominancia

En esta cláusula 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 Rec. UIT-R BT.601-5<sup>15</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  $C_R$ :

$$\underline{f}_{COHER\_COLOR} = (mean[C_B(i, j, t)], W_R \times mean[C_R(i, j, t)]): i, j, t \in \{\text{Región S-T}\},$$

y  $W_R = 1,5$

La anterior ecuación 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 Rec. UIT-R BT.601-5).

### D.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:

$$f_{CONT} = \{std[Y(i, j, t)]\}_P: i, j, t \in \{\text{Región S-T}\}$$

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

### D.7.5 Características basadas en información temporal absoluta (ATI, *absolute temporal information*)

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 información temporal absoluta,  $f_{ATI}$ , 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:

$$f_{ATI} = \{std |Y(i, j, t) - Y(i, j, t - 1)|\}_P: i, j, t \in \{\text{Región S-T}\}$$

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

---

<sup>15</sup> Las correcciones de la ganancia y del desplazamiento no se aplican a los planos de imagen  $C_B$  y  $C_R$ . Véase D.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 D.7.6) o para utilizarla en un modelo (véase la cláusula D.9), el requisito de un cuadro suplementario complica la tarea de emplazamiento de las regiones S-T (véase D.7.1.1).

### **D.7.6 Características basadas en el producto vectorial del contraste y la información temporal absoluta**

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 D.7.4 y D.7.5<sup>16</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.

## **D.8 Parámetros de calidad**

### **D.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 D.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 D.8.3), y la combinación de errores en el tiempo se denominará agrupamiento temporal (véase D.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 segundos. Los valores finales del parámetro agrupado en el tiempo pueden ajustarse y recortarse (véase D.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>16</sup> Se utiliza un producto vectorial estándar de las características  $f_{CONT}$  y  $f_{ATI}$  (es decir  $f_{CONT} \times f_{ATI}$ ) 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 D.8.2.1. No obstante, para las funciones de comparación *log\_loss* y *log\_gain* las características procesada y original se calculan como  $\log_{10}[f_{CONT}] \times \log_{10}[f_{ATI}]$ , y las funciones de comparación utilizan la resta (es decir  $f_p(s, t) - f_o(s, t)$  en vez de  $\log_{10}[f_p(s, t) / f_o(s, t)]$ ).

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

- 1) Comparar los valores originales de la característica con los valores de la característica procesados.
- 2) Ejecutar el agrupamiento espacial.
- 3) Ejecutar el agrupamiento temporal.
- 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.

## D.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 esta cláusula se presentan las funciones de enmascaramiento utilizadas por los diversos parámetros para producir parámetros de calidad en función del espacio y del tiempo.

### D.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:

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

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

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

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

siendo *pp* el operador parte positiva (es decir los valores negativos se sustituyen por cero) y *np* el operador parte negativa (es decir los valores positivos se sustituyen por cero).

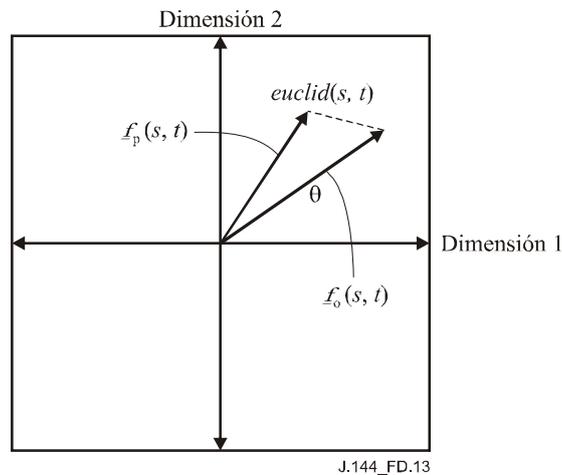
Estas funciones de enmascaramiento visual suponen que la percepción de la degradación es inversamente proporcional a la cantidad de actividad espacial o temporal localizada presente. Dicho de otro modo, las degradaciones espaciales resultan menos visibles conforme aumenta la actividad espacial (es decir 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.

### D.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,  $f_p(s, t)$ ,

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

La Figura D.13 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 D.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 Figura D.13 es la distancia euclidiana. La medida de la distancia euclidiana puede generalizarse a vectores de características que tengan un número arbitrario de dimensiones.



**Figura D.13 – Representación de la distancia euclidiana  $euclid(s, t)$  correspondiente a un vector de característica bidimensional**

### D.8.3 Funciones de agrupamiento espacial

Los parámetros de las regiones S-T (véase D.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 D.8.4. El Cuadro D.1 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  $s$  ([D-10]-[D-13]). 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  $log\_gain(s, t)$  del D.8.2.1 como el promedio del 5% más positivo de los valores sobre el índice

espacial  $s^{17}$ . 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  $ratio\_loss(s, t)$  del D.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).

#### D.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 D.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 segundos 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 segundos. 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 segundos 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 D.2 presenta un resumen de las funciones de agrupamiento temporal más utilizadas.

**Cuadro D.1 – Funciones de agrupamiento espacial y definiciones**

<b>Función de agrupamiento espacial</b>	<b>Definición</b>
<i>below5%</i> (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.
<i>above95%</i> (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.
<i>Mean</i> (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.
<i>std</i> (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.
<i>below5%tail</i> (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.

<sup>17</sup> Obsérvese que el índice temporal,  $t$ , no indica aquí cuadros individuales (véase D.7.1.1) Por contra, cada valor de  $t$  corresponde a regiones S-T con la misma amplitud temporal.

**Cuadro D.1 – Funciones de agrupamiento espacial y definiciones**

Función de agrupamiento espacial	Definición
<i>above99%tail</i> (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 D.2 – Funciones de agrupamiento temporal y sus definiciones**

Función de agrupamiento temporal	Definición
<i>10%</i>	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.
<i>25%</i>	Ordenar la historia temporal de los valores de parámetro de menor a mayor y seleccionar el nivel umbral del 25%.
<i>50%</i>	Ordenar la historia temporal de los valores de parámetro de menor a mayor y seleccionar el nivel umbral del 50%.
<i>90%</i>	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.
<i>Mean</i> (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.
<i>std</i> (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.
<i>above90%tail</i> (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.

### D.8.5 Recorte y ajuste no lineal

El parámetro  $p$  totalmente positivo o totalmente negativo agrupado temporalmente, véase D.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 (cláusula D.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*, y la función cuadrado, representada por *square*. Si se aplica la función *sqrt* 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 *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:

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

### D.8.6 Convenio de denominación de parámetros

En esta cláusula 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 D.3 resume los subnombres utilizados para crear el nombre técnico del parámetro, listados por orden de aparición. La cláusula D.8.6.1 contiene ejemplos de nombres técnicos de parámetros y sus subnombres asociados del cuadro D.3.

**Cuadro D.3 – 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	<i>Y</i> para el plano de luminancia de la imagen <i>color</i> para los planos de imagen ( <i>C<sub>B</sub></i> , <i>C<sub>R</sub></i> )
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 <i>HV</i> y <i>HV̄</i> . La relación subsiguiente de <i>HV</i> a <i>HV̄</i> se especifica mediante el nombre "Específico de la característica" (es decir en vez de ocupar un subnombre separado tras "Estadística del bloque").	<i>si13</i> para la característica <i>f<sub>SI13</sub></i> del D.7.2.2 <i>hv13_angleX.XXX_r_minYY</i> para la característica <i>f<sub>HV13</sub></i> del D.7.2.2, donde <i>X.XXX</i> es $\Delta\theta$ e <i>YY</i> es el <i>r<sub>min</sub></i> <i>coher_color</i> para la característica <i>f<sub>COHER_COLOR</sub></i> del D.7.3 <i>cont</i> para la característica <i>f<sub>CONT</sub></i> del D.7.4 <i>ati</i> para la característica <i>f<sub>ATI</sub></i> del D.7.5 <i>contrast_ati</i> para la característica <i>f<sub>CONTRAST_ATI</sub></i> del D.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.	<i>Deslizante</i>

**Cuadro D.3 – Convenio de denominación técnica utilizado para los parámetros de calidad de vídeo**

<b>Subnombre</b>	<b>Definición</b>	<b>Ejemplos</b>
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".	<i>Imagen</i>
Tamaño del bloque	Presente cuando la imagen está dividida en bloques S-T (véase D.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".	$8 \times 8$ para bloques que tengan 8 líneas de cuadro verticalmente por 8 píxeles de cuadro horizontalmente $128 \times 128$ 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 D.7.1.1), con referencia a vídeo de 30 cuadros por segundo (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" = $1 \times 1$ (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.	<i>mean</i> es el promedio de los valores de los píxeles  <i>std</i> es la desviación típica de los valores de los píxeles  <i>rms</i> 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 D.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.	<i>log_gain</i> (véase D.8.2.1) <i>ratio_loss</i> (véase D.8.2.1) <i>euclid</i> (véase D.8.2.2)

**Cuadro D.3 – Convenio de denominación técnica utilizado  
para los parámetros de calidad de vídeo**

Subnombre	Definición	Ejemplos
Función de agrupamiento espacial	Véase D.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 D.1
Función de agrupamiento temporal	Véase D.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 D.2
Función no lineal	Véase D.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 <i>sqrt</i> a un parámetro totalmente negativo, el parámetro se hará primero totalmente positivo (es decir se tomarán los valores absolutos).	<i>sqrt</i> para la raíz cuadrada del valor del parámetro temporalmente agrupado  <i>square</i> para el cuadrado del valor del parámetro agrupado temporalmente
Función de recorte	Véase D.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 <i>T</i> y cero por el nivel de recorte, y restar a continuación el nivel de recorte de todos los valores de parámetros resultantes.	<i>clip_0.45</i>  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.

### D.8.6.1 Ejemplos de nombres de parámetros

En esta cláusula se incluyen cinco ejemplos de nombres técnicos, y una descripción paso a paso del procedimiento de subdenominación del cuadro D.3.

*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 D.7.2.1 para preparar la extracción de la característica  $f_{SI13}$  del D.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 D.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 D.7.2.1 para preparar la extracción de la característica  $f_{HV13}$  del D.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 una  $\Delta\theta$  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 3,0 por 3,0. A continuación se calcula la característica  $f_{HV13}$  del D.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 D.7.6. El primer filtro, *contrast*, considerará los planos de luminancia directamente (véase D.7.4). El segundo filtro, *ati*, considerará imágenes generadas tomando las diferencias entre planos de luminancia correlativos (véase D.7.5). Las imágenes *contrast* y *ati* se tratan por separado hasta después de la fijación de umbrales.  $4 \times 4\_6F$  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 D.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 del D.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.

## **D.9 Modelo general**

En esta cláusula 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 D.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 Rec. UIT-R BT.601-5 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 D.7.2.2), dos se basan en características extraídas del vector formado por las dos componentes de crominancia ( $C_B$ ,  $C_R$ ) (véase D.7.3), y uno se basa en las características del

contraste y la información temporal absoluta, extraídos de la componente de luminancia  $Y$  (véanse D.7.4 y D.7.5, respectivamente).  $VQM_G$  viene dado por

$VQM_G =$

$$\begin{aligned} & \{-0,2097 \times Y_{si13\_8 \times 8\_6F\_std\_12\_ratio\_loss\_below5\%\_10\%} \\ & +0,5969 \times Y_{hv13\_angle0,225\_rmin20\_8 \times 8\_6F\_mean\_3\_ratio\_loss\_below5\%\_mean\_square\_clip\_0,06} \\ & +0,2483 \times Y_{hv13\_angle0,225\_rmin20\_8 \times 8\_6F\_mean\_3\_log\_gain\_above95\%\_mean} \\ & +0,0192 \times color\_coher\_color\_8 \times 8\_1F\_mean\_euclid\_std\_10\%\_clip\_0,6} \\ & -2,3416 \times [Y_{si13\_8 \times 8\_6F\_std\_8\_log\_gain\_mean\_mean\_clip\_0,004} |^{0,14}] \\ & +0,0431 \times Y_{contrast\_ati\_4 \times 4\_6F\_std\_3\_ratio\_gain\_mean\_10\%} \\ & +0,0076 \times color\_coher\_color\_8 \times 8\_1F\_mean\_euclid\_above99\%tail\_std} |_{0,0} \end{aligned}$$

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 una 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.

Si  $VQM_G > 1,0$ , entonces  $VQM_G = (1 + c) \times VQM_G / (c + VQM_G)$ , 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.

## D.10 Referencias informativas

- [D-1] Recomendación UIT-R BT.500-11 (2002), *Metodología para la evaluación subjetiva de la calidad de las imágenes de televisión.*
- [D-2] Recomendación UIT-T H.261 (1993), *Códec vídeo para servicios audiovisuales a  $p \times 64$  kbit/s.*
- [D-3] Recomendación UIT-T J.143 (2000), *Requisitos de usuario para mediciones objetivas de la percepción de la calidad vídeo en televisión digital por cable.*
- [D-4] Recomendación UIT-T P.910 (1999), *Métodos de evaluación subjetiva de la calidad vídeo para aplicaciones multimedios.*
- [D-5] Recomendación UIT-T P.931 (1998), *Retardo de las comunicaciones multimedios, sincronización y medición de la velocidad de tramas.*

- [D-6] Jain A.K., *Fundamentals of Digital Image Processing* (1989), Englewood Cliffs, NJ: Prentice-Hall Inc., pp. 348-357.
- [D-7] 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.
- [D-8] SMPTE 170M, *SMPTE Standard for Television – Composite Analog Video Signal – NTSC for Studio Applications*, Society of Motion Picture and Television Engineers, 595 West Hartsdale Avenue, White Plains, NY 10607.
- [D-9] SMPTE Recommended Practice 187 – 1995, *Center, Aspect Ratio, and Blanking of Video Images*, Society of Motion Picture and Television Engineers, 595 West Hartsdale Avenue, White Plains, NY 10607.
- [D-10] Wolf S. and Pinson M., *In-service performance metrics for MPEG-2 video systems*, in Proc. Made to Measure 98 – Measurement Techniques of the Digital Age Technical Seminar (1998), Conferencia técnica patrocinada conjuntamente por la International Academy of Broadcasting (IAB), la UIT y la Technical University of Braunschweig (TUB), Montreux, Suiza, 12-13 de noviembre de 1998.
- [D-11] Wolf S. and Pinson M., *Spatial-temporal distortion metrics for in-service quality monitoring of any digital video system* (1999), in Proc. SPIE International Symposium on Voice, Video, and Data Communications, Boston, MA, septiembre de 1999.
- [D-12] Wolf S. and Pinson M., *The relationship between performance and spatial-temporal region size for reduced-reference, in-service video quality monitoring systems* (2001), in Proc. SCI/ISAS 2001 (Systematics, Cybernetics, and Informatics/Information Systems Analysis and Synthesis), julio de 2001, pp. 323-328.
- [D-13] Wolf S. and Pinson M., *Video Quality Measurement Techniques* (2002), NTIA Report 02-392, junio de 2002.
- [D-14] Pinson M. and Wolf S., *Video Quality Measurement User's Manual* (2002), NTIA Handbook 02-1, febrero de 2002.
- [D-15] Material de referencia UIT-T (2004), *Objective perceptual assessment of video quality: Full reference television*.

#### **D.11 Datos objetivos brutos de la métrica de la calidad de vídeo (VQM)**

En esta cláusula se presentan todos los datos objetivos brutos de la VQM de la NTIA correspondientes a las pruebas VQEG fase 2 de televisión con referencia completa (FR-TV).

##### **Resumen de datos brutos**

Este modelo general desarrollado por la National Telecommunications and Information Administration (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 escala de calidad continua de doble estímulo (DSCQS). En los datos brutos presentados en este anexo 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 segundos 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 la cláusula D.6). Además, la región espacial de interés (SROI) utilizada para calcular el valor de la VQM de cada corto se escogió del siguiente modo:

- 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.
- 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 región válida procesada (que se abrevia por PVR, calculada automáticamente con arreglo a D.6.2) 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).
- 3) La SROI siempre se centra horizontalmente de modo que la muestra izquierda comience en un punto de muestreo de luminancia/crominancia de la Rec. UIT-R BT.601-5 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.

**Datos objetivos brutos para 525 líneas**

N.º Origen	N.º HRC	NTIA: Proponente H
1	1	0,660 (nota)
1	2	0,347
1	3	0,286
1	4	0,178
2	1	0,449
2	2	0,246
2	3	0,119
2	4	0,061
3	1	0,321
3	2	0,167
3	3	0,076
3	4	0,049
4	5	0,396
4	6	0,280
4	7	0,222
4	8	0,183
5	5	0,329
5	6	0,217
5	7	0,159
5	8	0,115
6	5	0,542
6	6	0,266
6	7	0,189
6	8	0,139
7	5	0,258
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

N.º Origen	N.º HRC	NTIA: Proponente H
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
9	14	0,124
10	9	0,666
10	10	0,250
10	11	0,375
10	12	0,129
10	13	0,078
10	14	0,153
11	9	0,513
11	10	0,534
11	11	0,407
11	12	0,161
11	13	0,148
11	14	0,159
12	9	0,600
12	10	0,410
12	11	0,471
12	12	0,244
12	13	0,171
12	14	0,114
13	9	0,537
13	10	0,425
13	11	0,346
13	12	0,215
13	13	0,188
13	14	0,169

NOTA – Para la fuente 1, HRC 1, el software 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 D.6.1.2. Para las otras escenas de HRC 1, el registro espacial/temporal se estimó correctamente. La cláusula D.6.1.5.7 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 software 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 software VQM habría registrado correctamente este corto y la puntuación objetiva sin procesar habría sido 0,529.

**Datos objetivos brutos para 625 líneas**

<b>N.º Origen</b>	<b>N.º HRC</b>	<b>NTIA: Proponente H</b>
1	2	0,421
1	3	0,431
1	4	0,264
1	6	0,205
1	8	0,155
1	10	0,123
2	2	0,449
2	3	0,473
2	4	0,312
2	6	0,260
2	8	0,226
2	10	0,145
3	2	0,472
3	3	0,506
3	4	0,308
3	6	0,239
3	8	0,183
3	10	0,146
4	2	0,409
4	3	0,458
4	4	0,384
4	6	0,354
4	8	0,280
4	10	0,232
5	2	0,470
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

<b>N.º Origen</b>	<b>N.º HRC</b>	<b>NTIA: Proponente H</b>
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
7	10	0,270
8	4	0,345
8	6	0,311
8	9	0,280
8	10	0,242
9	4	0,344
9	6	0,285
9	9	0,246
9	10	0,192
10	4	0,410
10	6	0,355
10	9	0,313
10	10	0,241
11	1	0,739
11	5	0,468
11	7	0,199
11	10	0,201
12	1	0,548
12	5	0,441
12	7	0,367
12	10	0,307
13	1	0,598
13	5	0,409
13	7	0,321
13	10	0,277

## **Apéndice I**

### **KDDI**

#### **Esquema de evaluación objetiva de la calidad de vídeo y valoración de la calidad de funcionamiento**

##### **I.1 Alcance**

En los últimos tiempos han empezado a ser de utilización habitual los servicios de radiodifusión y transmisión de televisión digital. En dichos servicios se utilizan códecs de vídeo (dispositivos de codificación de la señal de vídeo) basados en MPEG-2, un método normalizado internacional de compresión de señales de vídeo digitales. Los códecs de vídeo comprenden codificadores, que efectúan la compresión, y decodificadores, que reconstruyen los datos de vídeo comprimidos. Esos dispositivos funcionan eliminando información redundante del enorme volumen de información contenida en las señales de vídeo. Así es posible transmitir la información de manera eficaz utilizando sólo una porción limitada de anchura de banda.

Siempre se produce una cierta degradación de la calidad del vídeo que ha sido comprimido y transmitido utilizando un códec de vídeo. La importancia de la degradación depende del contenido de la imagen. Por lo general, se produce mayor distorsión en las escenas con movimiento rápido, por ejemplo, las de una transmisión deportiva. También hay variaciones en la calidad de la señal de salida generada por códecs diferentes. MPEG-2 es una norma internacional, pero la calidad de tipos específicos de vídeo comprimido sigue dependiendo, en cierta medida, de las características de la función códec implantada por el fabricante en su producto.

En la transmisión televisiva, sobre todo en la contribución, distribución primaria y distribución secundaria, hay que tratar de conseguir una alta calidad supervisando de manera continua la calidad de las imágenes transmitidas.

En la transmisión analógica convencional con modulación de frecuencia, se produce poca degradación de la imagen imputable al contenido o a la modulación analógica, con lo que la calidad permanece estable. Pero en la transmisión de vídeo digital comprimido, la calidad de la imagen varía, como se indica más arriba, de acuerdo con la naturaleza del contenido y el códec empleado y, previsiblemente, la comprobación de la calidad de este tipo de vídeo será una operación sumamente compleja.

Por ello, se considera necesario normalizar un esquema de evaluación de la calidad de imagen de los códecs de vídeo basados en la norma MPEG-2, utilizados principalmente en la contribución y las distribuciones primaria y secundaria. Se considera que en esas clases son necesarias las funciones siguientes:

- Evaluación genérica para diversos tipos de contenido de vídeo.  
Se soportan formatos de vídeo analógico/digital, compuesto/en componentes.
- Evaluación en tiempo real.
- Alineación precisa temporal y espacial entre una señal original y una señal de salida de códec.
- Evaluación sensible y exacta de distorsiones sutiles y complejas.

Teniendo en cuenta todo lo anterior, este apéndice propone un nuevo esquema de evaluación y su implementación en base a las características de la percepción visual humana, que permite mediciones muy precisas de la calidad vídeo.

## I.2 Esquema de evaluación de la calidad vídeo objetiva

La Figura I.1 muestra el modelo de evaluación de la calidad de imagen de tres capas según la visión del ojo humano. Por lo general, el ojo humano no puede ver un cuadro completo con una sola mirada, sino solamente una zona puntual de un cuadro, en torno al punto de observación del ojo humano, y reconoce la textura así como la calidad de la zona en función de los grados y características de la mezcla de ruido en esa textura. El cuadro completo se capta desplazando el punto de observación entre los objetos, que son componentes de imagen del cuadro, y al mismo tiempo se efectúa la evaluación de la calidad de imagen para la totalidad del cuadro. En este proceso, la calidad de imagen se determina por el ruido a lo largo de un cuadro. Por consiguiente, para efectuar una medición objetiva de la calidad de imagen subjetiva, se emplean estructuras de imagen de tres capas de macro a micro (capas objeto, textura y ruido), y se propone un esquema de ponderación del ruido de abajo a arriba que utiliza una función de ponderación particular en cada capa teniendo en cuenta la percepción visual humana (Figura I.2).

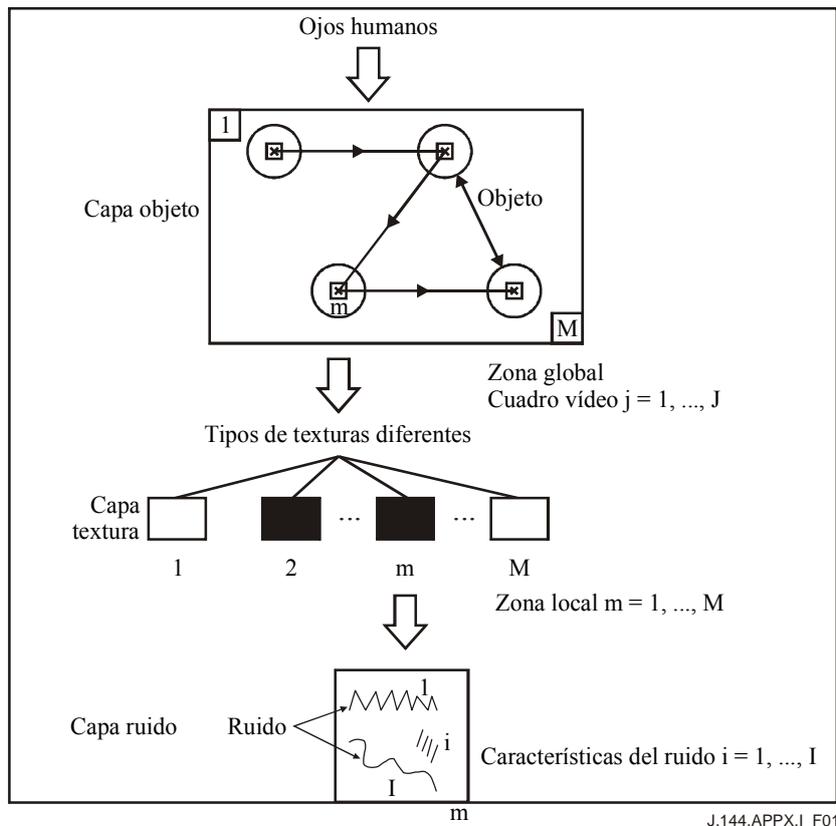
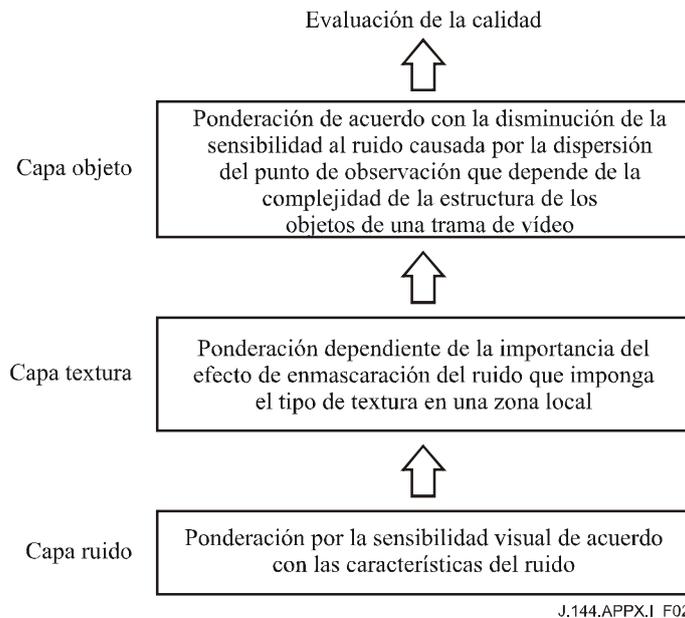


Figura I.1 – Modelo de tres capas para señal de vídeo



**Figura I.2 – Ponderación del ruido de abajo a arriba con estructura de tres capas**

En primer lugar, en la capa ruido, se ponderan los ruidos comunes de un proceso de compresión de vídeo, tales como el ruido de alta frecuencia, el ruido de baja frecuencia, el ruido de croma, la inestabilidad, el parpadeo y así sucesivamente, dependiendo de sus grados y características. Para esta ponderación, conviene efectuar una conversión de frecuencia a fin de clasificar estos ruidos. En segundo lugar, en la capa textura, se clasifican las zonas puntuales locales en varios grupos de acuerdo con sus tipos de textura. Dichos tipos incluyen, por ejemplo, "textura de detalle", tal como un bosque, árboles y un estadio en el que el ruido está fuertemente enmascarado, y "textura plana", tal como la piel de un ser humano, o bien el firmamento, en donde el ruido es fácilmente reconocible. Los ruidos se ponderan más o menos, por consiguiente, de acuerdo con sus tipos de textura. Por último, en la capa objeto, el grado de dispersión del punto de observación se predice midiendo el grado de complejidad de la estructura de los objetos en el cuadro de vídeo. A continuación se ponderan los ruidos de todo el cuadro en correspondencia con una disminución de la sensibilidad al ruido causada por la dispersión.

Para obtener las expresiones matemáticas de los procesos de ponderación, se define lo siguiente:

- $P(j,m,i)$ : Potencia de un ruido  $i$  en una zona local  $m$  de un cuadro  $j$ ;
- $h_i$ : Función de ponderación de un ruido  $i$ ;
- $C(j,m)$ : Textura en una zona local  $(j,m)$ ;
- $t_C$ : Función de ponderación del ruido en una textura  $C$ ;
- $G(j)$ : Parámetro que indica el grado de complejidad de la estructura de los objetos de un cuadro  $j$ ;
- $qG$ : Función de ponderación de un ruido dependiente del grado de dispersión de un punto de observación.

Teniendo en cuenta esas definiciones, los ruidos se suman en orden desde la capa inferior hasta la capa superior.

En la capa de ruido, sumando el ruido ponderado mediante la  $hi$  correspondiente a las características del ruido en una zona local  $(j,m)$ , se calcula  $WMSE_{NL}$  como sigue:

$$WMSE_{NL}(j,m) = \frac{1}{I} \sum_{i=1}^I hi \cdot P(j,m,i) \quad (I-1)$$

Seguidamente, en la capa de textura, sumando  $WMSE_{NL}(j,m)$  en toda la trama  $(m = 1, \dots, M)$  ponderado mediante la  $t_c$  correspondiente a una textura  $C(j,m)$  en una zona local  $(j,m)$ , se calcula  $WMSE_{TL}(j)$  como sigue:

$$WMSE_{TL}(j) = \frac{1}{M} \sum_{m=1}^M t_c(j,m) \cdot WMSE_{NL}(j,m) \quad (I-2)$$

Por último, en la capa objeto, tomando un valor medio de  $WMSE_{TL}$  en las tramas  $j = 1, \dots, J$  ponderadas mediante el  $G(j)$  correspondiente al grado de dispersión del punto de observación, se calcula  $WMSE_{OL}$  como sigue:

$$WMSE_{OL} = \frac{1}{J} \sum_{j=1}^J q_G(j) \cdot WMSE_{TL}(j) \quad (I-3)$$

A continuación se convierte  $WMSE_{OL}$  en  $WSNR$  y se calcula la escala de calidad continua de doble estímulo (DSCQS) (0-100%) definida en la Rec. UIT-R BT.500-11 como sigue:

$$WSNR(dB) = 10 \log_{10} \frac{255^2}{WMSE} \quad (I-4)$$

$$D(\%) = f(WSNR) \quad (I-5)$$

### Potencia del ruido en la zona local $P(j, m, i)$

Para empezar, se define la zona local  $m$  como un bloque cuadro de  $m_w \times m_h$ .

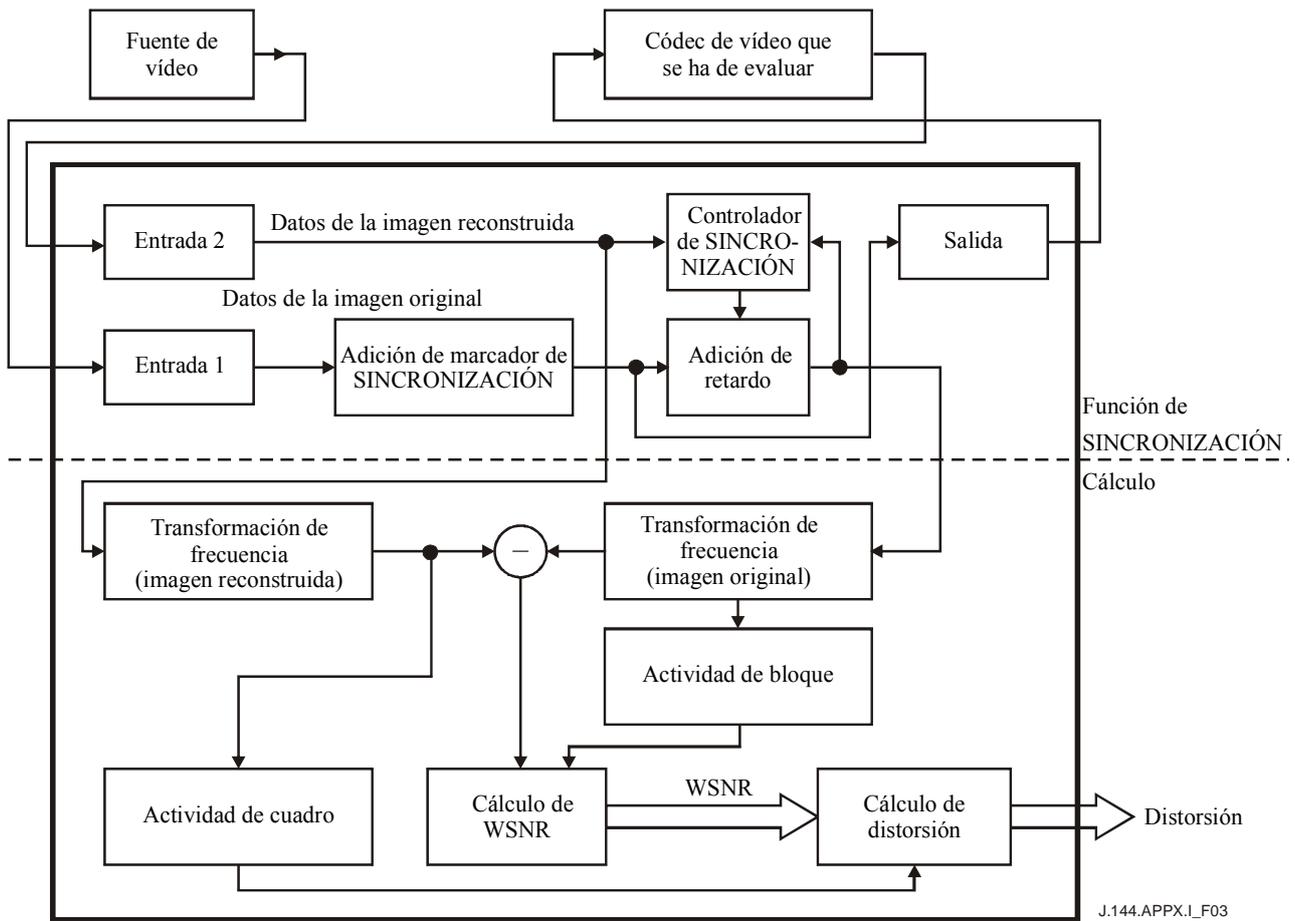
Sea la característica del ruido  $I = 1$  el ruido ponderado en el dominio de la frecuencia.

$$P(j,m,i) = \sum_{q=1}^{m_h} \sum_{p=1}^{m_w} \{X(p,q) - Y(p,q)\}^2$$

siendo  $X, Y$  el valor del coeficiente transformado de la imagen original y de la imagen codificada, respectivamente.

### I.3 Implementación

El sistema consta de dos partes: un módulo de sincronización, que permite efectuar una comparación precisa entre el vídeo reconstruido y el vídeo original, y un módulo de cálculo, que determina la calidad vídeo con referencia a las características de la percepción visual humana. La Figura I.3 muestra la configuración del sistema. En el Cuadro I.2 se describen los parámetros principales. Como muestra el Cuadro I.2, se soportan tanto señales compuestas (NTSC) como componentes con muestreos completos.



**Figura I.3 – Configuración del sistema**

### I.3.1 Módulo de sincronización

El sistema de evaluación de la calidad de vídeo en tiempo real necesita un módulo de sincronización, independiente del módulo de cálculo. Obsérvese que éste no es necesario para el cálculo fuera de línea, tal como ocurre en la comparación entre los ficheros de vídeo por el programa informático de evaluación de la calidad. A continuación se describe un ejemplo de método de sincronización.

Las señales de televisión procedentes de la fuente de vídeo original se introducen en el dispositivo del sistema mediante el módulo de entrada 1 y se marcan con un marcador de sincronización que varía con cada cuadro. El marcador es, por ejemplo, similar a una onda sinusoidal cuya frecuencia esté modulada por el número de cuadro. A continuación, los cuadros con marcadores se envían al módulo de retardo, en cuya memoria se almacenan. Al mismo tiempo, los cuadros se envían a través del módulo de salida al código de vídeo que se ha de evaluar. El código de vídeo comprime los cuadros, que se introducen de nuevo en el sistema a través del módulo 2 y se comparan con los cuadros marcados almacenados en el módulo de retardo del código de vídeo que se ha de evaluar. Por último, el módulo de sincronización lleva a cabo una alineación temporal (retardo de cuadro) y una alineación espacial (desplazamiento de línea y píxel) precisas, de modo que el grado de deterioro de la calidad que se describe más abajo concuerde lo más posible con la evaluación subjetiva efectuada por los observadores.

Con estas operaciones se obtiene la sincronización necesaria para la evaluación, y los marcadores utilizados en las mismas se diseñan de tal manera que funcionen correctamente incluso en los procesos con señal fuertemente distorsionada, tales como la compresión intensa, la separación Y/C y los filtrados en código de vídeo.

### I.3.2 Módulo de cálculo

A diferencia de lo que ocurre en la visión humana, el cálculo de la calidad de imagen tiene lugar siguiendo un proceso ascendente, construyendo el todo a partir de las diversas partes. En primer lugar, para evaluar el efecto de las variaciones de sensibilidad debidas a las frecuencias espaciales del ruido, se obtiene un ruido valor de la diferencia de los componentes de frecuencia de la imagen original y la imagen reconstruida. Ese valor se introduce en el módulo relación señal/ruido ponderada (WSNR, *weighted signal-to-noise ratio*), que asigna pesos de sensibilidad diferentes a cada región de frecuencia. Al mismo tiempo, obtiene un valor (la actividad de bloque) que indica si cada uno de los bloques de la imagen está vacío u ocupado. El efecto de enmascaramiento del ruido se aplica también para obtener una WSNR global.

Por último, se obtiene un valor que indica el tamaño de los objetos que forman la imagen (la actividad del cuadro). De esta manera, el sistema puede estimar el grado en que disminuye la sensibilidad al ruido debido a la dispersión de la degradación de la calidad, que se obtiene aplicando la disminución de la sensibilidad al ruido en la WSNR.

**Cuadro I.1 – Parámetros principales**

Formato de señal de vídeo aplicable	Señal compuesta NTSC Señal de componentes 525/60 Digital en serie D1
Frecuencia de muestreo (Entrada analógica)	14,318 MHz (NTSC) 13,5 MHz (componente Y) 6,75 MHz (componente C)
Códec aplicable	Códec basado en MPEG-1, 2 Códec compuesto, etc.
Zona de evaluación efectiva	768 píxels~480 líneas (NTSC) 720 píxels~480 líneas (componente Y) 360 píxels~480 líneas (componente C)
Análisis de señal	Transformada Hadamard (NTSC) Transformada de coseno discreta (componente) Alternativa: Transformada de Fourier
Ponderación de ruido	Sensibilidad visual a la frecuencia espacial Efecto de enmascaramiento del ruido Dispersión del punto de observación
Resultado de la evaluación	Evaluación de la calidad de la imagen (distorsión, %) WSNR (dB) SNR (dB)
Interfaz de la señal de control	RS-232C

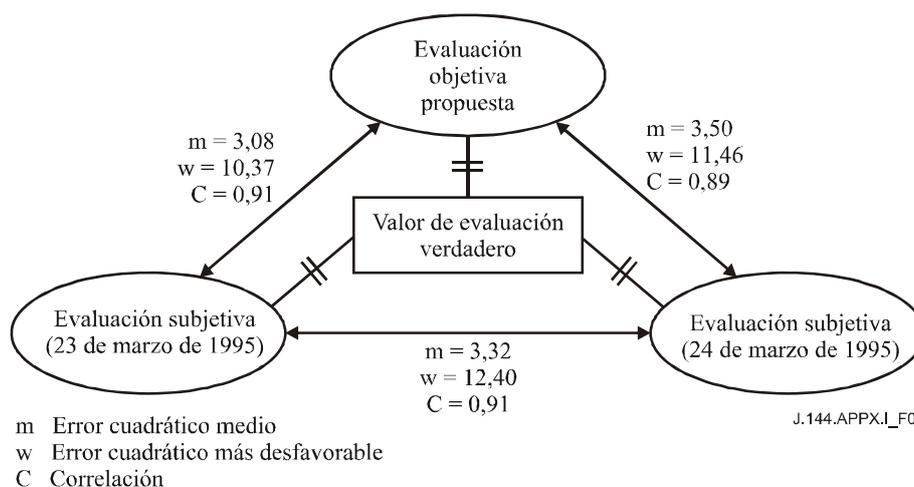
### I.4 Resultados de la verificación

Se compararon los resultados de la evaluación del esquema propuesto con los resultados de la prueba de evaluación subjetiva que ya habían sido calificados según Rec. UIT-R BT.500-11. Los objetivos de la evaluación son MPEG-2 SP@ML con 5 Mbit/s, 7 Mbit/s y 10 Mbit/s aplicados para las señales de prueba de televisión de componentes 4:2:2 de la Rec. UIT-R BT.601-5. Se trata de 17 datos entre los que se encuentran móvil, jardín con flores, animadoras, etc. Así pues se hay un total de 17 datos  $\times$  3 bits = 51 muestras (Cuadro I.2).

Con esas muestras se llevaron a cabo pruebas de evaluación subjetiva durante dos días (23 y 24 de marzo de 1995) en las mismas condiciones y con los mismos observadores. En la Figura I.4 se muestran el "triángulo" de resultados de la evaluación objetiva y de la evaluación subjetiva de dos días.

**Cuadro I.2 – Lista de datos de prueba**

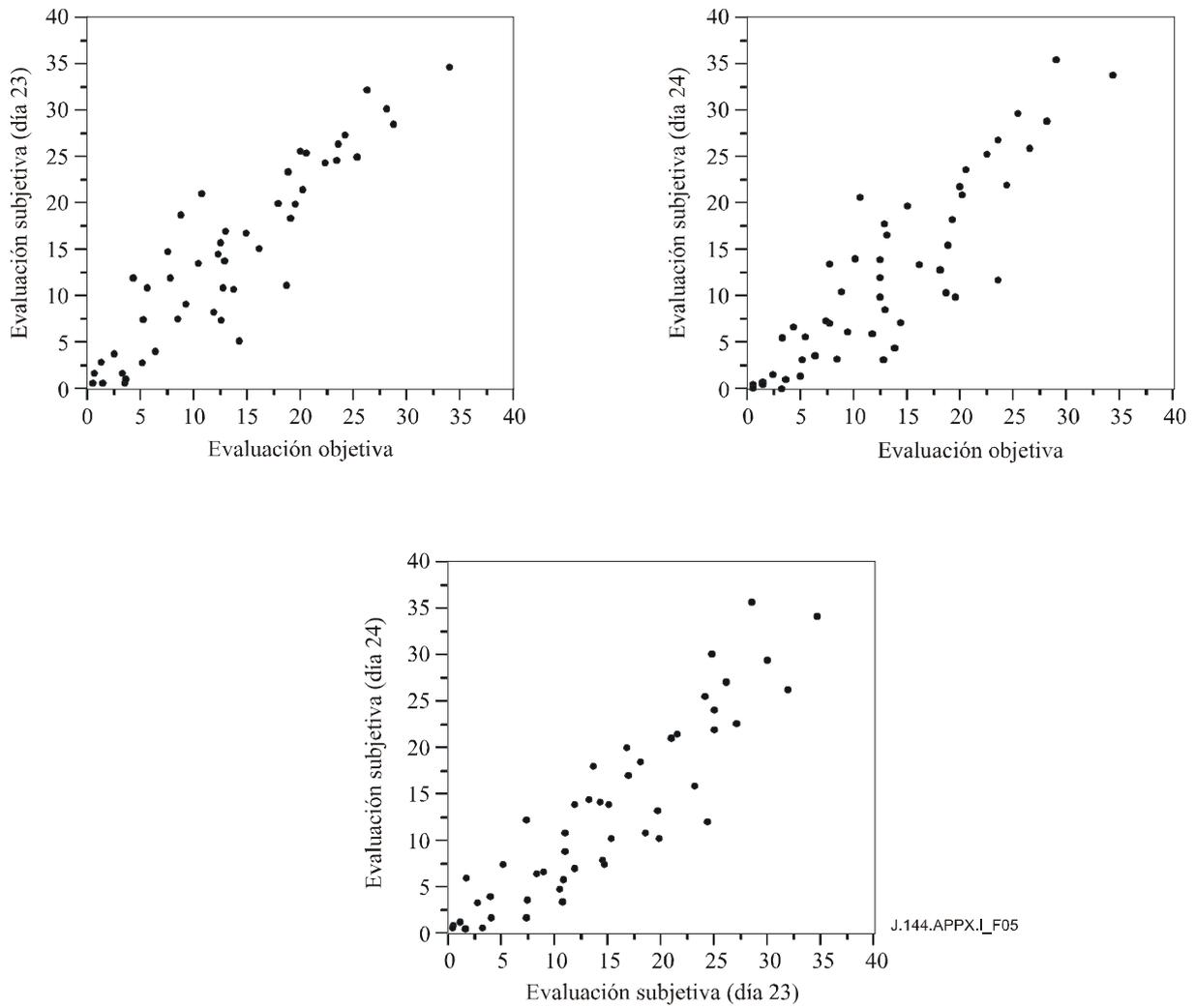
1	Susie
2	Amapola
3	Ping-Pong
4	Móvil y calendario
5	Hojas de otoño
6	Fútbol
7	Tormenta
8	Animadoras
9	Créditos
10	Navegación
11	Bicicleta
12	Equitación
13	Flores estivales
14	Noria
15	Jardín de flores
16	Puerto de Kiel 4
17	Ovillos



**Figura I.4 – Comparaciones con pruebas de evaluación subjetiva**

La Figura I.4 demuestra que la exactitud de la evaluación expresada en rmse, rwse y la correlación de los resultados de las tres evaluaciones es casi igual a partir del centro del triángulo, que es el valor verdadero de la evaluación. Además, la Figura I.5 presenta las distribuciones de las 51 muestras entre una evaluación objetiva y dos evaluaciones subjetivas. Las muestras de los tres gráficos están distribuidas aleatoriamente, pero puede apreciarse una diferencia sutil en cada distribución. En la distribución de la comparación subjetiva de los días 23 y 24, es uniformemente aleatoria, pero se pueden observar desigualdades en las distribuciones de las comparaciones de la evaluación objetiva y subjetiva dependiendo del rango de las notas de calificación. Esto es, tanto el gráfico del día 23 como el del día 24 correspondientes al esquema objetivo presentan trazados del muestreo con correlación más alta entre 20% y 40% pero menor correlación entre 10% y 20%. Se seguirá estudiando para eliminar este efecto.

De todo lo anterior se concluye que es posible utilizar el esquema propuesto además del de la Rec. UIT-R BT.500-11.



**Figura I.5 – Comparaciones entre una evaluación objetiva y dos evaluaciones subjetivas**

## Apéndice II

### Tektronix Inc. y Sarnoff Corporation

#### Medición objetiva de la calidad de percepción de vídeo utilizando una técnica de referencia completa basada en JND

##### II.1 Alcance, objeto y aplicación

##### II.1.1 Alcance

En este apéndice se define un método objetivo de medición de la calidad de vídeo basada en JND que utiliza la disponibilidad de la señal de vídeo con referencia completa. Se trata de un método de medición de doble extremo identificado como índice de calidad de imagen (PQR, *picture quality rating*) y representado en la Figura II.1.

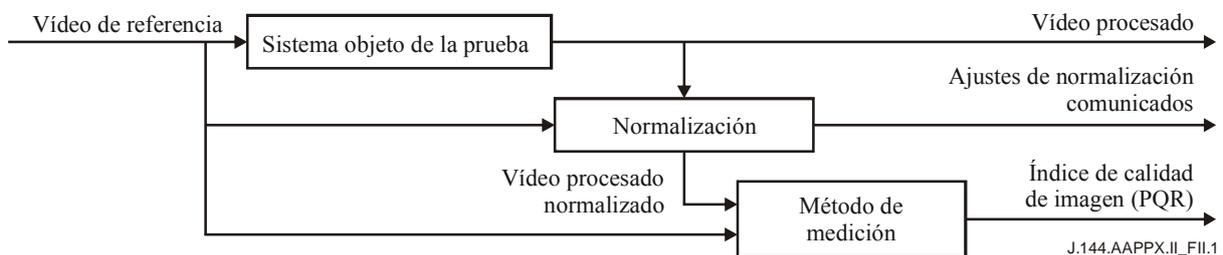


Figura II.1 – Diagrama de bloques del sistema

El método PQR especificado en el presente apéndice se basa en el procesamiento de vídeo de componente digital de 8 bits con arreglo a la definición de la Rec. UIT-R BT.601-5 de modo tal que sea representativo de la respuesta del sistema visual humano. Debido al carácter perceptivo de la medición pueden admitirse varios métodos de compresión (MPEG, NTSC, PAL, etc.). Además, el sistema de transmisión puede consistir en la concatenación de varios métodos de compresión o tener tan sólo un paso para la evaluación de un códec (combinación de codificador/decodificador). Los resultados del método PQR se expresan en valores del índice de calidad de imagen (PQR).

Para la aplicación del método PQR es necesario normalizar el vídeo procesado. En este apéndice sólo se especifican el algoritmo del método PQR y la precisión de la normalización. Los requisitos de la normalización pueden consultarse en II.3.1.

##### II.1.2 Objeto

Este apéndice contiene la descripción técnica de un método de medición de la calidad objetiva de vídeo perceptivo que está siendo utilizado actualmente. Aunque en el futuro pudieran desarrollarse métodos mejores, en este apéndice se presenta un método de medición de vídeo necesario para soportar la interconexión y el interfuncionamiento de las redes de telecomunicaciones en las interfaces con los sistemas de los usuarios finales, las portadoras, los proveedores de información y de servicios mejorados, y los equipos en las instalaciones del cliente.

## **II.1.3 Aplicación y limitaciones**

### **II.1.3.1 Aplicaciones del método PQR**

La aplicación de cualquier método de referencia completa conlleva ciertas restricciones operacionales. No obstante, el método PQR especificado en este apéndice no se limita a las evaluaciones en laboratorio. Entre las aplicaciones específicas para las que resulta adecuado utilizar el método PQR se encuentran:

- la evaluación, especificación y pruebas de aceptación de códecs,
- la supervisión de la transmisión en tiempo real y en servicio, en la fuente, y
- la evaluación de la transmisión remota con una copia de la referencia disponible.

Al utilizar el método PQR deben considerarse especialmente las limitaciones de precisión especificadas en II.1.3.4.

### **II.1.3.2 Limitaciones**

De acuerdo con la validación descrita en el Informe Final de la fase I del VQEG (material de referencia UIT-T) el método PQR especificado en el presente anexo es adecuado para secuencias de vídeo cortas (de 2 a 10 segundos de duración) a una distancia de observación de 5 h (5 alturas de imagen a 480 líneas por altura de imagen). Los valores PQR serán útiles para distancias de observación superiores o inferiores cuando la percepción humana de la degradación de la imagen disminuya al aumentar la distancia de observación, mientras que los valores PQR permanecen constantes. Aunque puede modificarse el algoritmo para reflejar la percepción humana a otras distancias de observación, estas modificaciones no forman parte del presente apéndice.

En el apéndice II – Adjunto 2 se presenta una lista detallada de factores de prueba, tecnologías de codificación y aplicaciones relativas a la precisión del método PQR, en base a los datos VQEG seleccionados.

Aunque el método de normalización especificado en II.3.1 puede utilizarse para detectar alteraciones en el tamaño de la imagen (tales como las producidas por una unidad de efectos especiales) no se ha podido obtener la información necesaria para determinar la magnitud del cambio de tamaño. El método PQR no es adecuado para la evaluación de imágenes que no sean del tamaño original de la imagen introducida en el sistema objeto de la prueba, ni para aquellas cuyos desplazamientos verticales consten de un número no entero de líneas.

Las clases de vídeo se definen en el anexo B/P.911. Este apéndice tiene por objeto presentar mediciones para las clases TV1, TV2 y TV3 definidas a continuación. La diferencia entre estas clases y las clases multimedios radica en que los codificadores siempre ofrecen una velocidad de cuadro constante y un funcionamiento de latencia constante. Aunque el sistema de compresión puede reducir el número de píxeles (normalmente sólo en dirección horizontal) como parte del proceso de codificación, la salida resultante del decodificador será vídeo de componentes de resolución completa con arreglo a la Rec. UIT-R BT.601-5.

- TV 0: sin pérdidas: Rec. UIT-R BT.601-5, 8 bits por muestra, vídeo utilizado para aplicaciones sin compresión.
- TV 1: utilizada para postproducción completa, muchas rectificaciones y capas de procesamiento, transmisión dentro de planta. Utilizada así mismo para la transmisión desde emplazamiento remoto a planta. Percepción transparente en comparación con TV 0.
- TV 2: utilizada para modificaciones sencillas, pocas rectificaciones, superposición de caracteres/logotipos, inserción de programas, y transmisión entre instalaciones. Un ejemplo de radiodifusión sería la transmisión desde la red a sus afiliados. Otros ejemplos serían el enlace descendente regional de un sistema de cable a una cabecera local y un sistema de videoconferencia de alta calidad. De percepción casi transparente en comparación con TV 0.

- TV 3: Utilizada para la difusión al hogar/consumidor (sin variaciones). Otros ejemplos son un sistema de cable desde la cabecera local a los hogares y la videoconferencia de calidad media a alta. Hay pocos artefactos en comparación con TV 2.

Todas estas clases tienen latencia unidireccional constante, aunque no necesariamente baja, y variación de retardo constante. El método PQR especificado en el presente apéndice no es adecuado para las aplicaciones de videoconferencia en la que se repiten cambios o que no satisfacen los requisitos de latencia y retardo de las clases de vídeo. Además, el método PQR sólo es aplicable a los sistemas de transmisión de radiodifusión típicos con tasas de errores muy bajas tales como los considerados en las pruebas VQEG.

### II.1.3.3 Comparación con la evaluación subjetiva

A pesar de la conveniencia de disponer de mediciones objetivas que mantengan una buena correlación con la evaluación subjetiva de calidad para alcanzar una calidad de servicio óptima, debe tenerse en cuenta que las mediciones objetivas no constituyen una sustitución directa de la evaluación subjetiva de la calidad. Las evaluaciones subjetivas de la calidad son procedimientos diseñados con el mayor esmero para determinar la opinión media de los observadores de un conjunto específico de secuencias de vídeo para una determinada aplicación. Los resultados de estas pruebas son útiles para el diseño básico del sistema y para la evaluación de referencias. Las evaluaciones subjetivas de calidad para una aplicación diferente con condiciones de prueba diferentes seguirán proporcionando resultados significativos. No obstante, es probable que las notas de opinión para el mismo conjunto de secuencias de vídeo tengan valores distintos. Las mediciones objetivas están destinadas a un amplio conjunto de aplicaciones produciendo los mismos resultados con un determinado conjunto de secuencias de vídeo. La selección de secuencias de vídeo a utilizar y la interpretación de las mediciones objetivas resultantes son algunos de los factores que variarán para cada aplicación específica. Por consiguiente las mediciones objetivas y la evaluación subjetiva de la calidad son complementarias y no intercambiables. Cuando la evaluación subjetiva sea adecuada por motivos relacionados con la investigación, se necesitarán mediciones objetivas para la especificación de los equipos y para la medición y supervisión diaria de la calidad de funcionamiento de los sistemas.

### II.1.3.4 Precisión y calibración cruzada

Para los valores PQR obtenidos de la ecuación II-36, se aplicaron los métodos de calibración cruzada y de precisión de VQM detallados en la Rec. UIT-T J.149 a los datos subjetivos de 525 líneas del VQEG, habiendo suprimido los HRC 15 y 16 (véase el material de referencia UIT-T).

Estos datos se seleccionaron por ser representativos de la televisión norteamericana. En la evaluación de la VQM, la curva de datos objetivos se ajusta a una escala normalizada (0-1) a fin de obtener la mejor correlación posible entre los datos subjetivos y objetivos. Esta transformación también es importante para el cálculo de la precisión, ya que proporciona la mayor coherencia de confianza estadística del poder de resolución (precisión) en el rango de valores objetivos.

Se utiliza la función de ajuste de curvas denominada Logística II en 4.2/J.149, con la restricción de interceptar o aproximarse asintóticamente a cero para imágenes cada vez más perfectas ( $a = -e^{-cd}$ ).

$$VQM = a + \frac{b - a}{1 + e^{-c(PQR-d)}}$$

La función logística está sometida asimismo a la restricción de tener un valor máximo de 1 ( $b = 1$ ) en la escala normalizada 0–1 y de 100 en la escala DSCQS nativa (véase la referencia UIT-R BT.500-11) que se utiliza para las pruebas subjetivas de la VQM con referencia completa. Una vez transformado a la escala común, el método PQR puede someterse a calibración cruzada con cualquier otra VQM. Con estas restricciones, la ecuación de la función logística para el método PQR queda como se indica a continuación.

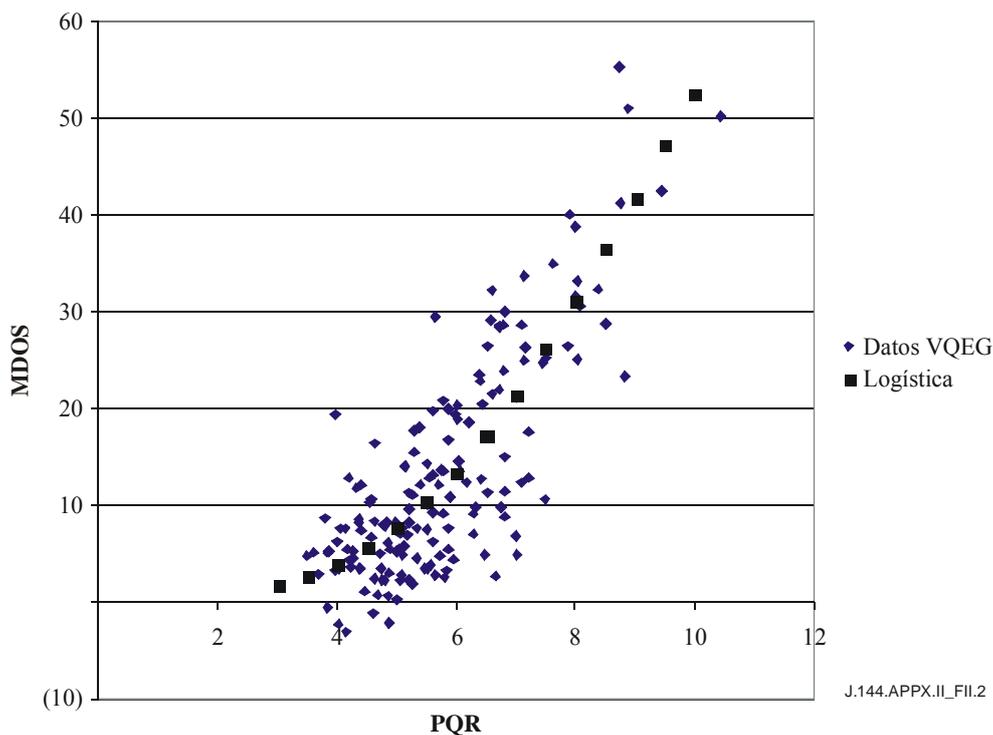
$$VQM = \frac{1 - e^{-c \times PQR}}{1 + e^{c(d - PQR)}}$$

Siendo:

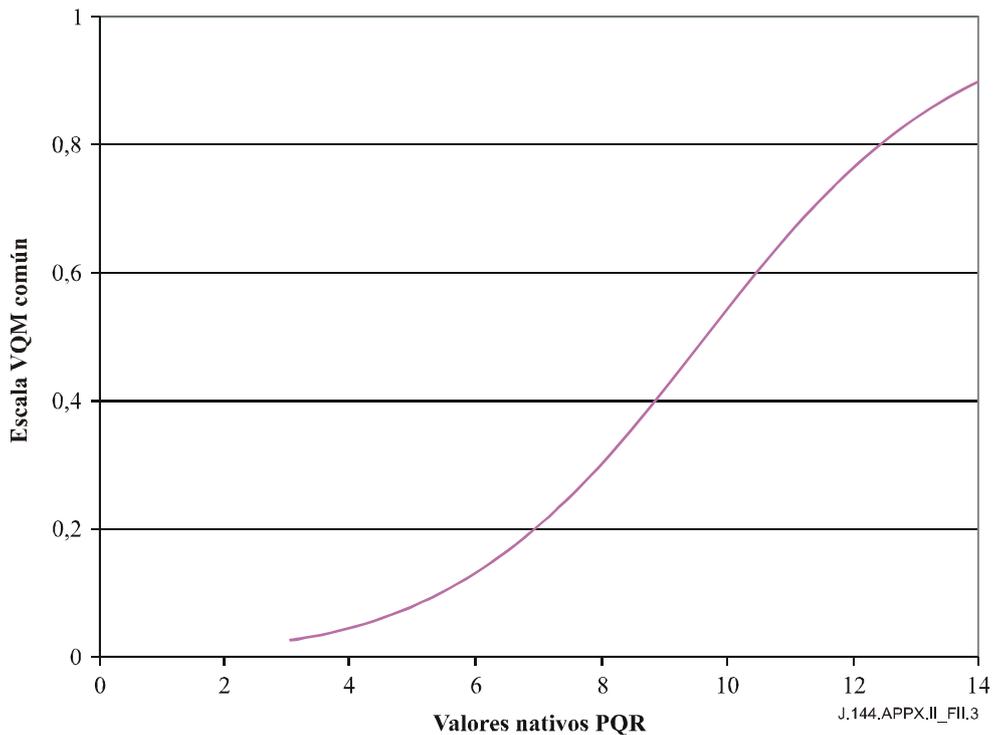
$$c = 0,5031$$

$$d = 9,634$$

La Figura II.2 muestra la gráfica de los datos VQEG seleccionados y la función logística restringida para los valores PQR nativos. La Figura II.3 muestra la conversión de los valores PQR nativos a la escala común utilizando la función logística restringida. Aunque la curva de la función logística restringida puede calcularse para extenderse más allá de los puntos de los datos VQEG llegando a alcanzar límites adecuados para los cálculos del valor PQR y las notas DSCQS de datos subjetivos, sólo se considera válida para el análisis de precisión en el rango de valores PQR nativos entre 3,5 y 10,5.

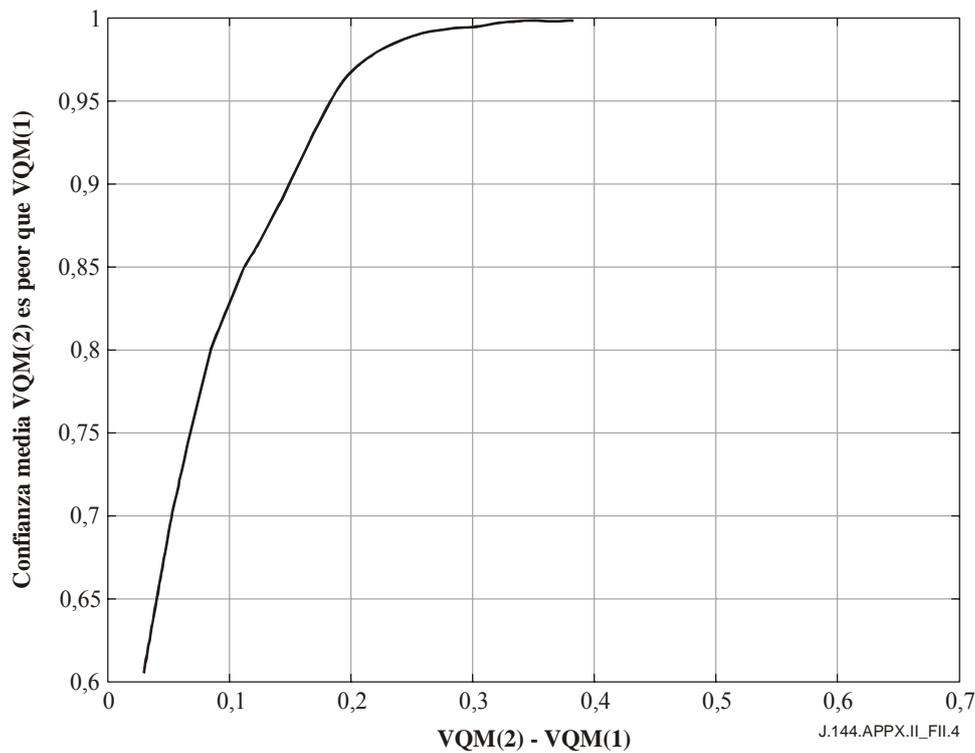


**Figura II.2 – Datos VQEG y función logística restringida**



**Figura II.3 – Conversión de los valores PQR de la escala nativa a la común**

El poder de resolución es la diferencia entre los valores VQM de las dos mediciones de modo que haya un nivel de confianza conocido de que la medición con el mejor valor VQM tenga también la mejor nota subjetiva. En la Rec. UIT-T J.149 se exponen dos métodos para calcular el poder de resolución de una VQM. Hay un complejo método estadístico denominado la prueba z que considera que el poder de resolución no es tan sólo función de la confianza requerida sino que varía con arreglo a los puntos de datos subjetivos/objetivos disponibles. La Figura II.4 muestra el resultado de dicho cálculo para el método PQR. Los valores VQM delta se expresan en el dominio normalizado mientras que el poder de resolución es nominalmente constante para todos los valores VQM.



**Figura II.4 – Poder de resolución en función de la escala VQM normalizada**

La gráfica muestra que para una diferencia de 0,1 en valores PQR normalizados hay un nivel de confianza de 0,81 de que la secuencia con el menor valor PQR (la mayor calidad) tenga la nota subjetiva DSCQS más baja (mayor calidad). La elección del poder de resolución a utilizar en las diversas aplicaciones se facilita por la clasificación de un análisis de errores como se muestra en el anexo C.

Un cálculo de primer orden del poder de resolución puede realizarse simplemente a partir del error cuadrático medio (RMSE, *root mean square error*) de las notas subjetivas en función de los valores subjetivos en el dominio normalizado. Diferencias en los valores VQM iguales al RMSE proporcionan un nivel de confianza del 68%, mientras que 1,96 veces el RMSE proporciona un nivel de confianza del 95%. Aunque con este método no se obtienen los mismos resultados que con el planteamiento más complejo, resulta fácil de entender y puede ser de gran utilidad considerando los niveles de precisión en los entornos operacionales.

$$VQM\_RMSE = 0,06723$$

Esto corresponde aproximadamente a la curva más precisa de la Figura II.4 como se indica a continuación.

Nivel de confianza	Figura II.4	RMSE
68%	0,053	0,066
95%	0,187	0,132

La Figura II.4 puede convertirse a la escala PQR, si se desea, ajustando a escala el eje x (es decir,  $VQM(2) - VQM(1)$ ) utilizando la derivada de la función de ajuste de curvas logística II mostrada en 4.2/J.149.

$$[PQR(2) - PQR(1)] = [VQM(2) - VQM(1)] \frac{\left(1 + e^{-c(PQR-d)}\right)^2}{c(b-a)e^{-c(PQR-d)}}$$

Aplicando las restricciones  $a = -e^{-cd}$  y  $b = 0$  la ecuación se convierte en

$$[PQR(2) - PQR(1)] = [VQM(2) - VQM(1)] \frac{(1 + e^{c(d-PQR)})^2}{c(e^{c(d-PQR)} + e^{-c(PQR)})}$$

Siendo:

$$c = 0,5031$$

$$d = 9,634$$

Esto genera una familia de curvas ya que el factor de escala del eje x depende del valor PQR.

## II.2 Referencias

Las siguientes normas contienen disposiciones que, por ser mencionadas en este texto, constituyen disposiciones del presente apéndice. Al efectuar esta publicación estaban en vigor las ediciones indicadas. Todas las normas son susceptibles de revisión por lo que se invita a las partes en los acuerdos basados en el presente apéndice a considerar la posibilidad de aplicar las ediciones más recientes de las normas indicadas a continuación.

- Recomendación UIT-R BT.601-5 (1995), *Parámetros de codificación de televisión digital para estudios con formatos de imagen normal 4:3 y de pantalla ancha 16:9.*
- Recomendación UIT-T J.149 (2004), *Método para especificar la precisión y la calibración cruzada de métricas de la calidad de vídeo.*
- Recomendación UIT-T P.911 (1998), *Métodos de evaluación subjetiva de la calidad audiovisual para aplicaciones multimedios.*

## II.3 Introducción

### II.3.1 Normalización

Normalizar significa suprimir los cambios sistemáticos invariables en el tiempo, del vídeo de la entrada de referencia al vídeo de salida, antes de llevar a cabo la medición basada en el modelo de visión humana. (Véase la Figura II.1.) El método PQR se basa en filtros del modelo de visión humana que comparan las imágenes de referencia y las procesadas exactamente píxel a píxel. El método de normalización especificado en T1.TR.73-2001\*, anexo B (véase el apéndice II.A) es adecuado para utilizarlo con el método PQR.

Los parámetros que han de ajustarse en el proceso de normalización son el desplazamiento horizontal y vertical de la imagen; los cambios de ganancia del color y de luminancia; los cambios de nivel de c.c. del color y de luminancia; y la diferencia de retardo canal a canal entre componentes o entre luminancia y color. Dado que estos cambios podrían provocar otros en la calidad de la imagen percibida, se informará de ellos junto con los resultados de la prueba. Es necesario separar estos cambios del cálculo del PQR por dos motivos. El principal es obtener el valor de PQR más preciso posible. En segundo lugar, esta normalización corresponde estrechamente al funcionamiento típico del sistema en relación con los parámetros del nivel de c.c. y la ganancia, donde los oportunos ajustes suelen ser viables y efectuarse con normalidad. Se considera que pequeños valores de desplazamiento de la imagen, ya sea horizontal o verticalmente no suelen modificar la calidad de la imagen percibida. No obstante, su presencia constituye un error de la imagen y generará importantes problemas en aplicaciones multigeneración. La alineación temporal debe ser perfecta para que cada cuadro/campo procesado se compare con la referencia equivalente.

---

\* Las normas T1 son mantenidas por ATIS desde noviembre 2003.

El vídeo procesado se normaliza campo a campo por medición de las señales de prueba calibradas, insertadas en la frecuencia de referencia. Sólo se suprimen del vídeo los cambios estáticos invariables en el tiempo, los cambios dinámicos debidos a los procesos de compresión y descompresión se miden como parte del cálculo del PQR. La normalización del vídeo procesado antes de los cálculos del PQR deberá satisfacer las tolerancias del Cuadro II.1. Los valores del PQR basados en normalización que no satisfagan las tolerancias del Cuadro II.1 tendrán una precisión inferior a la especificada en el apartado II.1.3.4.

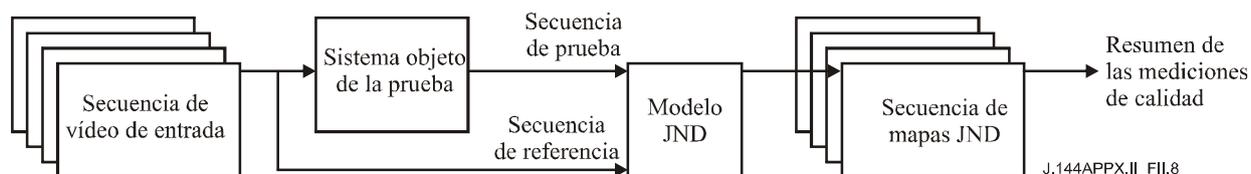
**Cuadro II.1 – Parámetros y tolerancias de normalización**

Parámetro	Tolerancia de normalización
Nivel de luminancia	< 0,2 dB del blanco de cresta
Nivel de diferencia de color	< 0,2 dB de la máxima excursión permitida
Nivel de c.c. de luminancia	< 0,5 % del blanco de cresta
Nivel de c.c. de la diferencia de color	< 0,5% de la máxima excursión permitida
Diferencia de retardo entre canales	< 2 ns
Desplazamiento de los píxeles horizontales	< 0,1 píxel
Desplazamiento de las líneas verticales	0 líneas (limitado a un número entero de desplazamientos de línea)
Desplazamiento temporal	0 campos

### II.3.2 Presentación del método de medición PQR

El método PQR predice los índices perceptivos que los observadores asignarán a una secuencia de imágenes de color degradada en relación con su homóloga no degradada. Al modelo se le introducen dos secuencias de imágenes que producen varias estimaciones de las diferencias, entre ellas una sola métrica de las diferencias perceptivas entre las secuencias. Estas diferencias se cuantifican en unidades de diferencia apenas perceptible (JND, *just-noticeable difference*) del modelo humano.

Una secuencia de vídeo de entrada atraviesa dos canales distintos en dirección al observador (no representado en la figura). Uno de los canales está intacto (el canal de referencia), mientras que el otro distorsiona la imagen de algún modo (el canal objeto de la prueba). La distorsión, efecto colateral de alguna de las medidas adoptadas por razones de economía (tal como la compresión), puede tener lugar en un codificador antes de la transmisión, en el propio canal de transmisión, o bien en el proceso de decodificación. En la Figura II.8, el bloque denominado "Sistema objeto de la prueba" se refiere esquemáticamente a cualquiera de estas alternativas. El método PQR se implementa sustituyendo la pantalla y el observador por el modelo de visión humana, que compara las secuencias de prueba y de referencia a fin de producir una secuencia de mapas JND en vez de la evaluación subjetiva.

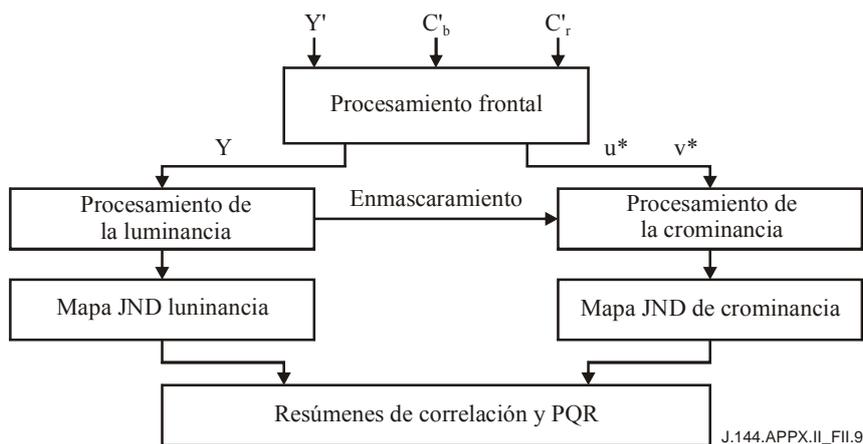


**Figura II.8 – Modelo de visión humana en la evaluación de un sistema<sup>18</sup>**

<sup>18</sup> Las Figuras II.5 a II.7 se han omitido a propósito.

La Figura II.9 representa de forma esquemática el algoritmo. Las entradas son dos secuencias de imágenes de longitud arbitraria. Para cada campo de cada secuencia de entrada hay tres conjuntos de datos representados por  $Y'$ ,  $C'_b$ , y  $C'_r$  en la parte superior de la Figura II.9 obtenidos, por ejemplo, de una cinta D1. A continuación los datos  $Y$ ,  $C_b$ ,  $C_r$  se transforman en tensiones catódicas  $R'$ ,  $G'$ ,  $B'$  que producen los valores de los píxeles visualizados. En el modelo, las tensiones  $R'$ ,  $G'$  y  $B'$  se someten a otro proceso para transformarlas en una imagen de luminancia y dos cromáticas que se pasan a las etapas subsiguientes.

El objeto del procesamiento frontal es transformar las señales de entrada de vídeo en salidas luminosas para transformar después estas salidas luminosas en cantidades definidas psicofísicamente que caractericen por separado la luminancia y la crominancia.



**Figura II.9 – Diagrama de bloques del modelo de visión humana**

La etapa de procesamiento de luminancia acepta dos imágenes (la de prueba y la de referencia) de luminancia  $Y$ , expresadas como fracciones de la máxima luminancia de la pantalla. A partir de estas entradas, la etapa de procesamiento de la luminancia genera un mapa JND de luminancia. Este mapa consiste en una imagen cuyos niveles de gris son proporcionales al número de JND entre las imágenes de prueba y de referencia en la posición del píxel correspondiente.

Un procesamiento similar, basado en el espacio de color uniforme CIE  $L^*u^*v^*$ , tiene lugar para cada una de las imágenes de crominancia  $u^*$  y  $v^*$ . Las salidas del procesamiento de  $u^*$  y  $v^*$  se combinan para generar el mapa JND de crominancia. Tanto el procesamiento de la crominancia como el de la luminancia sufren la influencia de las entradas procedentes del canal de luminancia que se denomina enmascaramiento, y que hace que las diferencias percibidas sean visibles en mayor o menor medida, dependiendo de la estructura de las imágenes de luminancia.

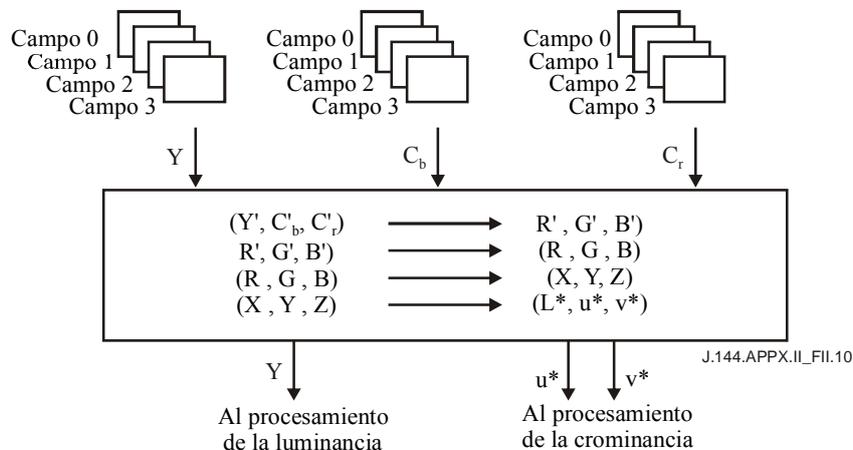
Los mapas JND de luminancia, de crominancia y el combinado de luminancia y crominancia, están disponibles a la salida, junto con un pequeño número de resúmenes de mediciones obtenidas de estos mapas. Los resúmenes de valores PQR sencillos modelan el índice global que daría un observador a las distorsiones de una secuencia de prueba. Los mapas JND proporcionan una visión más detallada de la posición y gravedad de los artefactos.

## II.4 Presentación del algoritmo

### II.4.1 Procesamiento frontal

La pila de los cuatro campos denominados  $Y'$ ,  $C'_b$ ,  $C'_r$  en la parte superior de la Figura II.10 representa el conjunto de cuatro campos consecutivos procedentes de una prueba o de una secuencia de imágenes de referencia. En la primera etapa de procesamiento se transforman los datos  $Y'$ ,  $C'_b$ ,  $C'_r$  en las tensiones catódicas  $R'$ ,  $G'$ ,  $B'$  (véase II.5.1.1).

En la segunda etapa de procesamiento, se aplica a cada imagen  $R'$ ,  $G'$  y  $B'$  una no linealidad puntual. En esta etapa se realiza la transferencia de las tensiones catódicas  $R'$ ,  $G'$  y  $B'$  a intensidades de pantalla ( $R$ ,  $G$  y  $B$ ) (fracciones de la luminancia máxima) en el modelo. La no linealidad recorta además las luminancias pequeñas en cada plano de la pantalla.



**Figura II.10 – Esquema de la primera etapa de procesamiento**

Tras la no linealidad, existen dos opciones de procesamiento: la de media altura y la de altura completa. En la exploración entrelazada, las imágenes de media altura<sup>19</sup> se procesan conforme están, sin líneas intermedias en blanco. El modelado de altura completa está disponible para exploraciones progresivas (en las que un campo contiene un cuadro, es decir una sola imagen en vez de dos campos entrelazados).

A continuación, el vector  $(R,G,B)$  de cada píxel del campo se somete a una transformación lineal (que depende de los fósforos de la pantalla) pasando a coordenadas  $(X, Y, Z)$  triestímulo CIE 1931. El componente de luminancia  $Y$  de este vector se somete a procesamiento de luminancia.

Para aproximarse (en cada píxel) a la uniformidad perceptiva del espacio de color con diferencias de color isoluminantes, los píxeles individuales se convierten en CIELUV, norma internacional de espacio de color uniforme. Los componentes de crominancia  $u^*$ ,  $v^*$  de este espacio se someten a los pasos de procesamiento de crominancia del modelo.<sup>20</sup>

## II.4.2 Procesamiento de la luminancia

Como se observa en la Figura II.11, cada valor de luminancia se somete en primer lugar a una no linealidad compresiva. A continuación, cada campo de luminancia se filtra y se reduce su densidad de muestreo con arreglo a una pirámide gaussiana de cuatro niveles, a fin de modelar la descomposición observada psicofísicamente y fisiológicamente de las señales visuales de entrada en distintas bandas de frecuencias espaciales. Tras esta descomposición, el grueso del procesamiento subsiguiente del modelo consiste en operaciones semejantes (por ejemplo filtrado orientado) que se ejecuta en cada uno de los niveles de la pirámide.

<sup>19</sup> Las filas en una imagen de media altura corresponden a un campo, es decir a las líneas pares, o bien a las impares, de un cuadro.

<sup>20</sup> El canal de luminancia  $L^*$  de CIELUV no se utiliza en el procesamiento de luminancia, sino que se sustituye por una no linealidad visual para la que se ha calibrado el modelo de visión en un rango de valores de luminancia.  $L^*$  se utiliza en el procesamiento de crominancia, no obstante, para crear una métrica de crominancia que sea aproximadamente uniforme y conocida por los diseñadores de las pantallas.

Una vez completado el proceso de construcción de la pirámide, la imagen de menor resolución de la pirámide se somete a filtrado temporal y a cálculo de contraste, mientras que los otros tres niveles se someten a filtrado espacial y a cálculo de contraste. En cada caso el contraste es la diferencia local de valores de píxel dividida por una suma local, convenientemente ajustada a escala. Inicialmente, esto establece la definición de 1 JND, que se pasa a las etapas subsiguientes del modelo.<sup>21</sup> (La calibración revisa iterativamente las interpretaciones de 1-JND en las etapas intermedias del modelo.) El valor absoluto de la respuesta de contraste se pasa a la etapa siguiente, conservándose el signo algebraico para reinsertarlo justo antes de la comparación de las imágenes (cálculo del mapa JND).

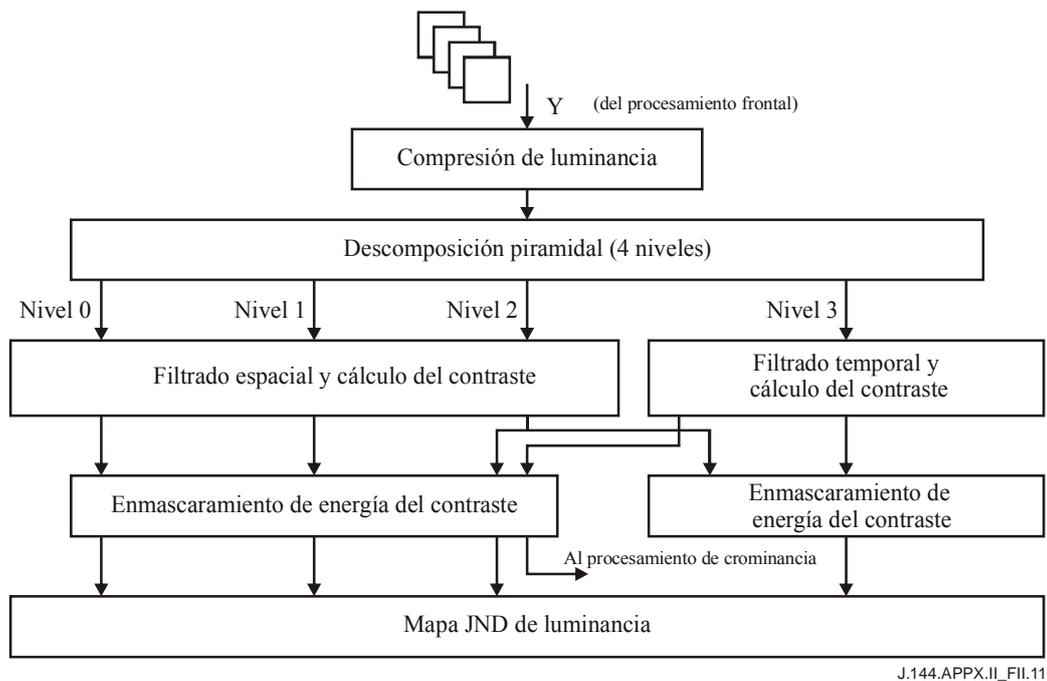
La siguiente etapa (enmascaramiento del contraste) es una operación de ajuste de ganancia en la que la respuesta de contraste se divide por una función de todas las respuestas de contraste. Esta atenuación combinada de cada respuesta por las otras respuestas locales se incluye para modelar efectos de "enmascaramiento" visuales tales como la disminución de la sensibilidad a las distorsiones en las zonas de imagen "ocupada". En esta etapa del modelo, se construye una estructura temporal (parpadeo) para enmascarar las diferencias espaciales, así como una estructura espacial para enmascarar las diferencias temporales. El enmascaramiento de luminancia se aplica también en el lado de la crominancia, como se expone a continuación.

Las respuestas de contraste enmascaradas (junto con los signos de los contrastes) se utilizan para generar el mapa JND de luminancia. Esto se lleva a cabo:

- separando cada imagen en sus componentes positiva y negativa (rectificación de media onda);
- creando agrupaciones locales (promediando y reduciendo la densidad de muestreo, para modelar la agregación espacial local observada en los experimentos psicofísicos);
- evaluando las diferencias absolutas de imagen canal por canal;
- aumentando la densidad de muestreo hasta la misma resolución (que será la mitad de la resolución de la imagen original debido a la etapa de agrupamiento);
- evaluando la norma Q de Minkowski en todos los canales.

---

<sup>21</sup> La asociación de un contraste constante a 1 JND es una aplicación de la denominada ley de Weber para la visión.



J.144.APPX.II\_FIL.11

**Figura II.11 – Esquema del procesamiento de luminancia**

### II.4.3 Procesamiento de la crominancia

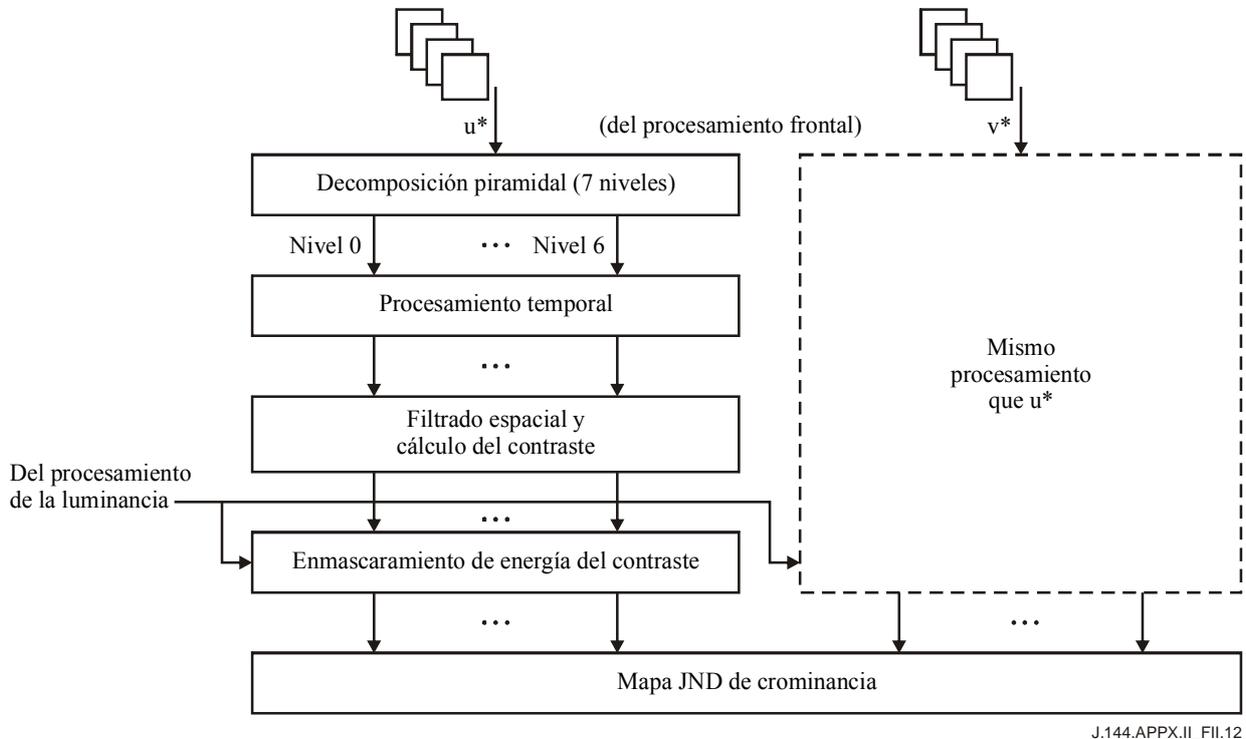
El procesamiento de la crominancia se asemeja al de la luminancia en varios aspectos. Para definir los umbrales de detección correspondientes al modelo de crominancia se utilizan las diferencias de crominancia ( $u^*$  y  $v^*$ ) del espacio CIELUV dentro de la propia imagen; análogamente a como se utiliza el contraste (y la ley de Weber) para definir el umbral de detección en el modelo de luminancia. Además, como en el modelo de luminancia, los "contrastes" cromáticos definidos por las diferencias de  $u^*$  y  $v^*$  se someten a una etapa de enmascaramiento. Una no linealidad de transductor hace que la discriminación de un incremento de contraste entre una imagen y otra dependa de la respuesta de contraste común a ambas imágenes.

La Figura II.12 muestra que, como en el procesamiento de la luminancia, cada componente de crominancia  $u^*$ ,  $v^*$  se somete a una descomposición piramidal. No obstante, mientras que el procesamiento de la luminancia sólo necesita cuatro niveles piramidales, el de la crominancia utiliza siete. Esto obedece al hecho empírico de que los canales cromáticos son sensibles a frecuencias espaciales muy inferiores a los de los canales de luminancia. Asimismo, tiene en cuenta el factor intuitivo de que las diferencias de color pueden observarse en regiones grandes y uniformes.

Para reflejar esta insensibilidad intrínseca al parpadeo de los canales de crominancia, se ejecuta el procesamiento temporal promediándose sobre cuatro campos de imagen.

A continuación, se efectúa el filtrado espacial en  $u^*$  y  $v^*$  mediante un núcleo de Laplace. Esta operación produce una diferencia de color en  $u^*$ ,  $v^*$  que (por definición de espacio de color uniforme) está conectada métricamente con diferencias de color apenas perceptibles. El valor 1 en esta etapa significa que se ha conseguido una sola JND, que desempeña un papel análogo al del contraste del canal de luminancia basado en la ley de Weber. (Como en el caso de la luminancia, la unidad de crominancia 1 JND debe volver a interpretarse durante la calibración.)

Este valor de diferencia de color se pondera, y su valor absoluto se pasa (con el signo algebraico del contraste) a la etapa de enmascaramiento de contraste. La etapa de enmascaramiento ejecuta la misma función que en el modelo de luminancia, aunque es algo más sencilla, porque sólo recibe la entrada de los canales de luminancia y crominancia cuya diferencia se está evaluando. Por último, las respuestas de contraste enmascarado se procesan exactamente como en el modelo de luminancia (véase el último párrafo del II.4.2).



J.144.APPX.II\_FII.12

**Figura II.12 – Esquema del procesamiento de la crominancia**

#### II.4.4 Resúmenes de salida

Para cada campo de la secuencia de vídeo comparada, se combinan en primer lugar los mapas JND de luminancia y crominancia para generar un mapa JND total. Este mapa JND total se calcula como raíz cuadrada de la suma de los cuadrados de los valores del mapa de luminancia y crominancia, píxel a píxel.

A continuación, cada uno de los tres mapas JND (luminancia, crominancia y luminancia y crominancia combinadas) se reduce a un índice de un solo número denominado valor índice de calidad de imagen (PQR, *picture quality rating*). Estos índices de un solo número se calculan mediante la norma Q de Minkowski. De acuerdo con este planteamiento, cada valor de píxel del mapa JND se eleva a la Q-ésima potencia. A continuación se calcula el PQR como la raíz Q-ésima de la suma normalizada de todos los valores de píxel elevados a la Q-ésima potencia.

A continuación, se calculan sólo tres medidas de la calidad de funcionamiento para muchos campos de una secuencia de vídeo (una para la luminancia, otra para la crominancia y una tercera para la combinación de luminancia y crominancia). Los valores PQR correspondientes a cada campo de la secuencia se reducen a un índice de calidad de imagen para toda la secuencia, también con arreglo a la norma Q de Minkowski.

Aunque el método PQR es válido para un rango de longitudes de secuencias, en este apéndice los valores PQR deberán calcularse para 60 cuadros (2 segundos) de vídeo. Es importante asimismo tener en cuenta que los valores PQR para secuencias de medio segundo o menos pueden resultar difíciles de comparar con la evaluación subjetiva. Esto se debe al hecho de que los datos de evaluación subjetiva carecen de fiabilidad para secuencias tan cortas.

## II.5 Detalles del algoritmo

La aplicación del modelo de medición del PQR descrito en este apéndice requiere utilizar valores paramétricos que calibren el algoritmo a fin de aproximar la respuesta al sistema visual humano. El Cuadro II.2 muestra los valores paramétricos que deberán utilizarse en la implementación de este apéndice.

**Cuadro II.2 – Valor de los parámetros del modelo PQR**

Tipo de parámetro	Símbolo	Valor	Cláusula
Compresión de luminancia	m	0,65	II.5.2.1
	$L_d$	7,5 cd/m <sup>2</sup>	
Filtrado temporal de 60 Hz	$t_e$	33/64	II.5.2.3.2
	$t_l$	31/64	
Filtrado temporal de 50 Hz	$t_e$	11/16	II.5.2.3.2
	$t_l$	5/16	
Umbral de contraste de luminancia (por nivel piramidal)	$w_0$	1/150	II.5.2.4
	$w_1$	1/900	
	$w_2$	1/1280	
	$w_3$	1/500	
Contrastes de enmascaramiento de luminancia	$\beta$	1,4	II.5.2.5
	a	3/32	
	c	5/32	
	$m_f$	10/1024	
	$m_t$	50	
	$m_{ft}$	3/64	
Umbral de contraste de crominancia (por nivel piramidal)	$q_0$	384	II.5.3.4
	$q_1$	60	
	$q_2$	24	
	$q_3$	6	
	$q_4$	4	
	$q_5$	3	
	$q_6$	3	

**Cuadro II.2 – Valor de los parámetros del modelo PQR**

Tipo de parámetro	Símbolo	Valor	Cláusula
Constantes de enmascaramiento de crominancia	$\beta_c$	1,4	II.5.3.5
	$a_c$	0,5	
	$c_c$	0,5	
	$m_c$	10/1024	
	$k$	0,7	

### II.5.1 Procesamiento frontal

En la Figura II.10, el procesamiento frontal transforma las señales de entrada de vídeo  $Y'$ ,  $C'_b$  y  $C'_r$  en tensiones catódicas en primer lugar, en valores de luminancia de los tres fósforos a continuación y, por último, en variables psicofisiológicas que se separan en componentes de luminancia y crominancia. El valor triestímulo  $Y$ , calculado en II.5.1.3, sustituye el "valor de intensidad del modelo" utilizado antes de añadir el procesamiento de crominancia al modelo JND. Además, se generan las componentes de crominancia  $u^*$  y  $v^*$ , píxel a píxel, con arreglo a las especificaciones de color uniforme CIE.

#### II.5.1.1 De ( $Y'$ , $C'_b$ , $C'_r$ ) a ( $R'$ , $G'$ , $B'$ )

Los pasos bosquejados a continuación describen la transformación de los cuadros de imagen  $Y'$ ,  $C'_b$ ,  $C'_r$  en las tensiones  $R'$ ,  $G'$ ,  $B'$  que excitan la pantalla. En este contexto, los apóstrofes indican que las señales de entrada se han sometido previamente a corrección gamma en el codificador. Estas señales, a su vez transformadas, excitan un dispositivo de pantalla CRT<sup>22</sup> cuya función de transferencia voltaje-corriente puede aproximarse con gran exactitud por una no linealidad gamma.

Se supone aquí que las imágenes digitales de entrada se encuentran en el formato 4:2:2, o sea: resolución completa en la correlacionada de luminancia  $Y'$  y media resolución horizontalmente en las correlacionadas de crominancia  $C'_b$  y  $C'_r$ . Se supone que los datos  $Y'$ ,  $C'_b$ ,  $C'_r$  se guardan en el orden especificado en la Rec. UIT-R BT.601-5, o sea,

$$C'_{b0}, Y'_0, C'_{r0}, Y'_1, C'_{b1}, Y'_2, C'_{r1}, Y'_3, \dots, C'_{bn/2-1}, Y'_{n-1}, C'_{rn/2-1}, Y'_{n-2}, \dots$$

##### *Paso 1*

Introducir las matrices  $Y'$ ,  $C'_b$ ,  $C'_r$  de un solo cuadro. A continuación expandir las matrices  $C'_b$  y  $C'_r$  hasta alcanzar la resolución completa de la imagen  $Y'$ . Las matrices  $C'_b$  y  $C'_r$  se tienen inicialmente media resolución en sentido horizontal, por lo que hay que aumentar su densidad de muestreo para crear los campos de resolución completa. Para comenzar, los píxeles alternativos  $C'_b$ ,  $C'_r$  de una fila se asignan a la  $Y'_i$  de numeración par a la que flanquean en el tren de datos. A continuación, se calcula el par  $C'_b$ ,  $C'_r$  que ha de asociarse a la  $Y'_i$  de numeración impar promediándolas con sus vecinas horizontales inmediatas.

##### *Paso 2*

Empaquetar las matrices de resolución completa  $Y'$ ,  $C'_b$ ,  $C'_r$  en dos campos. En el caso de  $Y'$ , el primer campo contiene las líneas impares de esta matriz, mientras que el segundo contiene sus líneas pares. Se ejecuta el mismo procesamiento sobre las matrices  $C'_b$  y  $C'_r$  para generar los dos primeros campos  $C'_b$  y  $C'_r$ .

<sup>22</sup> En II.5.1.2.1 puede consultarse la descripción del modelo de pantalla CRT.

### Paso 3

Para cada píxel de cada uno de los dos campos, convertir los correspondientes valores  $Y'$   $C'_b$   $C'_r$  en los valores catódicos de entrada  $R'$ ,  $G'$ ,  $B'$ . A los efectos del presente apéndice, los valores  $Y'$   $C'_b$   $C'_r$  están relacionados con los valores  $R'G'B'$  por la siguiente ecuación:

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1.371 \\ 1 & -0,336 & -0,698 \\ 1 & 1.732 & 0 \end{bmatrix} \begin{bmatrix} Y' \\ C'_b \\ C'_r \end{bmatrix} \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \quad (\text{II-1})$$

Las matrices  $R'$ ,  $G'$  y  $B'$  ya pueden someterse a la siguiente etapa del algoritmo de procesamiento frontal.

#### II.5.1.2 De $(R', G', B')$ a $(R, G, B)$

##### II.5.1.2.1 Transformación del valor del píxel

Calcular para cada píxel la fracción de luminancia máxima  $R$  correspondiente a la entrada  $R'$ . Análogamente, calcular las fracciones de luminancia  $G$  y  $B$  de las entradas  $G'$ ,  $B'$ . La máxima luminancia de cada cátodo se supone corresponde al valor de entrada 255. Las ecuaciones siguientes describen la transformación de  $(R', G', B')$  en  $(R, G, B)$ :

$$\begin{aligned} R &= \left[ \frac{\text{máx}(R', t_d)}{255} \right]^\gamma \\ G &= \left[ \frac{\text{máx}(G', t_d)}{255} \right]^\gamma \\ B &= \left[ \frac{\text{máx}(B', t_d)}{255} \right]^\gamma \end{aligned} \quad (\text{II-2})$$

En este contexto, se supone que el valor umbral por defecto  $t_d$  es igual a 16, que corresponde al nivel de negro de la pantalla, y que el de  $\gamma$  es de 2,5 por defecto. Se ha escogido el valor 16 para  $t_d$  a fin de que la pantalla tenga un margen dinámico de 1000:1 aproximadamente (es decir,  $(255/16)^{2.5}$ ).

##### II.5.1.2.2 Opciones de procesamiento de imagen de altura completa y de media altura

El modelo PQR tiene dos opciones de especificación para la representación vertical de imágenes  $(R, G, B)$ : para cada cuadro (en las imágenes progresivas) y para los campos pares e impares (en las imágenes entrelazadas).

1) *Cuadro*

Las imágenes son de altura completa y contienen una imagen explorada progresivamente.

2) *Entrelazado de media altura*

Las imágenes de media altura se procesan directamente.

Las primeras seis subcláusulas de II.5.2 y II.5.3 describen el procesamiento de la luminancia y de la crominancia de altura completa. Las cláusulas II.5.2.7 y II.5.3.7 describen el procesamiento de media altura.

### II.5.1.3 De (R, G, B) a (X, Y, Z)

Calcular los valores triestímulo CIE 1931 X, Y y Z para cada píxel a partir de valores de luminancia fraccionaria R, G, B. Esto supone que las entradas, que dependen del dispositivo de pantalla, sean las siguientes:  $(x_r, y_r)$ ,  $(x_g, y_g)$  y  $(x_b, y_b)$  las coordenadas de cromaticidad de los tres fósforos, y  $(x_w, y_w)$  la cromaticidad del blanco de referencia del monitor.

Las cromaticidades de blanco de referencia  $(x_w, y_w) = (0,3127, 0,3290)$  corresponden al iluminante D65. El Cuadro II.3 muestra las opciones para las coordenadas de los fósforos de la pantalla.

**Cuadro II.3 – Opciones para las coordenadas de los fósforos de la pantalla**

Origen	$(x_r, y_r)$	$(x_g, y_g)$	$(x_b, y_b)$
UIT-R BT.709-5 (SMPTE 274M)	(0,640; 0,330)	(0,300; 0,600)	(0,150; 0,060)
SMPTE 240M	(0,630; 0,340)	(0,310; 0,595)	(0,155; 0,070)
EBU	(0,640; 0,330)	(0,290; 0,600)	(0,150; 0,060)

Dados los anteriores valores de parámetros, los valores X, Y, Z del píxel se obtienen a partir de las siguientes ecuaciones:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \frac{x_r}{y_r} Y_{or} & \frac{x_g}{y_g} Y_{og} & \frac{x_b}{y_b} Y_{ob} \\ Y_{or} & Y_{og} & Y_{ob} \\ \frac{z_r}{y_r} Y_{or} & \frac{z_g}{y_g} Y_{og} & \frac{z_b}{y_b} Y_{ob} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (\text{II-3})$$

Donde,  $z_r = (1-x_r-y_r)$ ,  $z_g = (1-x_g-y_g)$ ,  $z_b = (1-x_b-y_b)$ , y los valores  $Y_{or}$ ,  $Y_{og}$ ,  $Y_{ob}$  vienen dados por la siguiente ecuación

$$\begin{bmatrix} Y_{or} \\ Y_{og} \\ Y_{ob} \end{bmatrix} = \begin{bmatrix} \frac{x_r}{y_r} & \frac{x_g}{y_g} & \frac{x_b}{y_b} \\ 1 & 1 & 1 \\ \frac{z_r}{y_r} & \frac{z_g}{y_g} & \frac{z_b}{y_b} \end{bmatrix}^{-1} \begin{bmatrix} \frac{x_w}{y_w} \\ 1 \\ \frac{z_w}{y_w} \end{bmatrix} \quad (\text{II-4})$$

siendo  $z_w = (1-x_w-y_w)$ .

También se necesitarán los valores triestímulo  $X_n$ ,  $Y_n$ ,  $Z_n$  del blanco de referencia del dispositivo. Estos valores corresponden a la cromaticidad  $(x_w, y_w)$  y son tales que, a la máxima activación del fósforo ( $R' = G' = B' = 255$ ),  $Y = 1$ . Los valores triestímulo correspondientes al blanco de referencia son  $(X_n, Y_n, Z_n) = (x_w/y_w, 1, z_w/y_w)$ .

Como paso final de la obtención de los valores X, Y, Z, se realiza un ajuste para acomodar un supuesto de luz ambiente y compensar el reflejo de la pantalla de visualización. Este ajuste tiene la forma siguiente:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \leftarrow \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \left( \frac{L_a}{L_{\text{máx}}} \right) \begin{bmatrix} X_n \\ Y_n \\ Z_n \end{bmatrix} \quad (\text{II-5})$$

Ahora, se introducen y asignan dos parámetros que puede especificar el usuario,  $L_{\text{máx}}$  y  $L_a$ , como valores por defecto.  $L_{\text{máx}}$ , máxima luminancia de la pantalla, se hace igual a  $100 \text{ cd/m}^2$  como corresponde a las pantallas comerciales. La luminancia compensadora del deslumbramiento,  $L_a$ , se hace igual a  $5 \text{ cd/m}^2$ , en congruencia con los valores de pantalla medidos en condiciones de la Rec. UIT-R BT.500-11.

Se supone que la cromaticidad de la luz ambiente coincide con la del blanco de referencia de la pantalla. Hay que tener en cuenta, no obstante, que en la opción del modelo de luminancia exclusiva, en la que no se calcula el punto neutro  $(X_n, Y_n, Z_n)$ , se efectúa el ajuste

$$Y \leftarrow Y + \frac{L_a}{L_{\text{máx}}} \quad (\text{II-6})$$

en vez de aceptar el resultado de la ecuación II-5. Esto equivale a utilizar la componente Y de la ecuación II-5 debido a que  $Y_n$  es siempre igual a 1. Obsérvese asimismo que la cantidad  $L_{\text{máx}} * Y$  es la luminancia de la pantalla en  $\text{cd/m}^2$ .

#### II.5.1.4 De (X, Y, Z) a (L\*, u\*, v\*)

Los valores X, Y, Z, se transforman píxel a píxel, al sistema de color uniforme 1976 CIELUV:

$$L^* = 116 \left( \frac{Y}{Y_n} \right)^{1/3} - 16 \quad \text{para } \frac{Y}{Y_n} > 0,008856 \quad (\text{II-7})$$

$$L^* = 903,3 \left( \frac{Y}{Y_n} \right) \quad \text{para } \frac{Y}{Y_n} \leq 0,008856$$

$$u^* = 13L^* (u' - u'_n) \quad (\text{II-8})$$

$$v^* = 13L^* (v' - v'_n) \quad (\text{II-9})$$

siendo,

$$u' = \frac{4X}{(X + 15Y + 3Z)} \quad (\text{II-10})$$

$$v' = \frac{9Y}{(X + 15Y + 3Z)} \quad (\text{II-11})$$

$$u'_n = \frac{4X_n}{(X_n + 15Y_n + 3Z_n)} \quad (\text{II-12})$$

$$v'_n = \frac{9Y_n}{(X_n + 15Y_n + 3Z_n)} \quad (\text{II-13})$$

Obsérvese que la coordenada  $L^*$  no se tiene en cuenta en los cálculos de la luminancia.  $L^*$  sólo se utiliza en el cálculo de las coordenadas de crominancia  $u^*$  y  $v^*$ <sup>23</sup>. Por consiguiente, de las cifras anteriores, sólo se conservan para su procesamiento posterior las imágenes  $u^*$  y  $v^*$ .

## II.5.2 Procesamiento de luminancia

Véase la Figura II.13. En esta cláusula, las imágenes de la prueba de entrada y del campo de referencia se representan por  $I_k$  e  $I_k^{\text{ref}}$  ( $k = 0, 1, 2, 3$ ). Los valores de los píxeles de  $I_k$  e  $I_k^{\text{ref}}$  se representan por  $I_k(i,j)$  e  $I_k^{\text{ref}}(i,j)$ , respectivamente. Se empieza por los valores triestímulo  $Y$  calculados en el procesamiento frontal. En el texto siguiente sólo se consideran los campos  $I_k$ . El procesamiento de  $I_k^{\text{ref}}$  es idéntico.  $k=3$  representa el campo más reciente de una secuencia de 4 campos.

Las cláusulas II.5.2.1 a II.5.2.6 describen el procesamiento de altura completa. La cláusula II.5.2.7 trata de las modificaciones necesarias para el procesamiento de media altura.

### II.5.2.1 Compresión de luminancia

El primer paso del modelo de luminancia consiste en una no linealidad integrada por una función de potencia descendente desplazada en una constante. Sea  $Y_3(i,j)$ , la matriz de luminancia relativa del último campo, donde 3 representa el último campo. Se cumplirá entonces que:

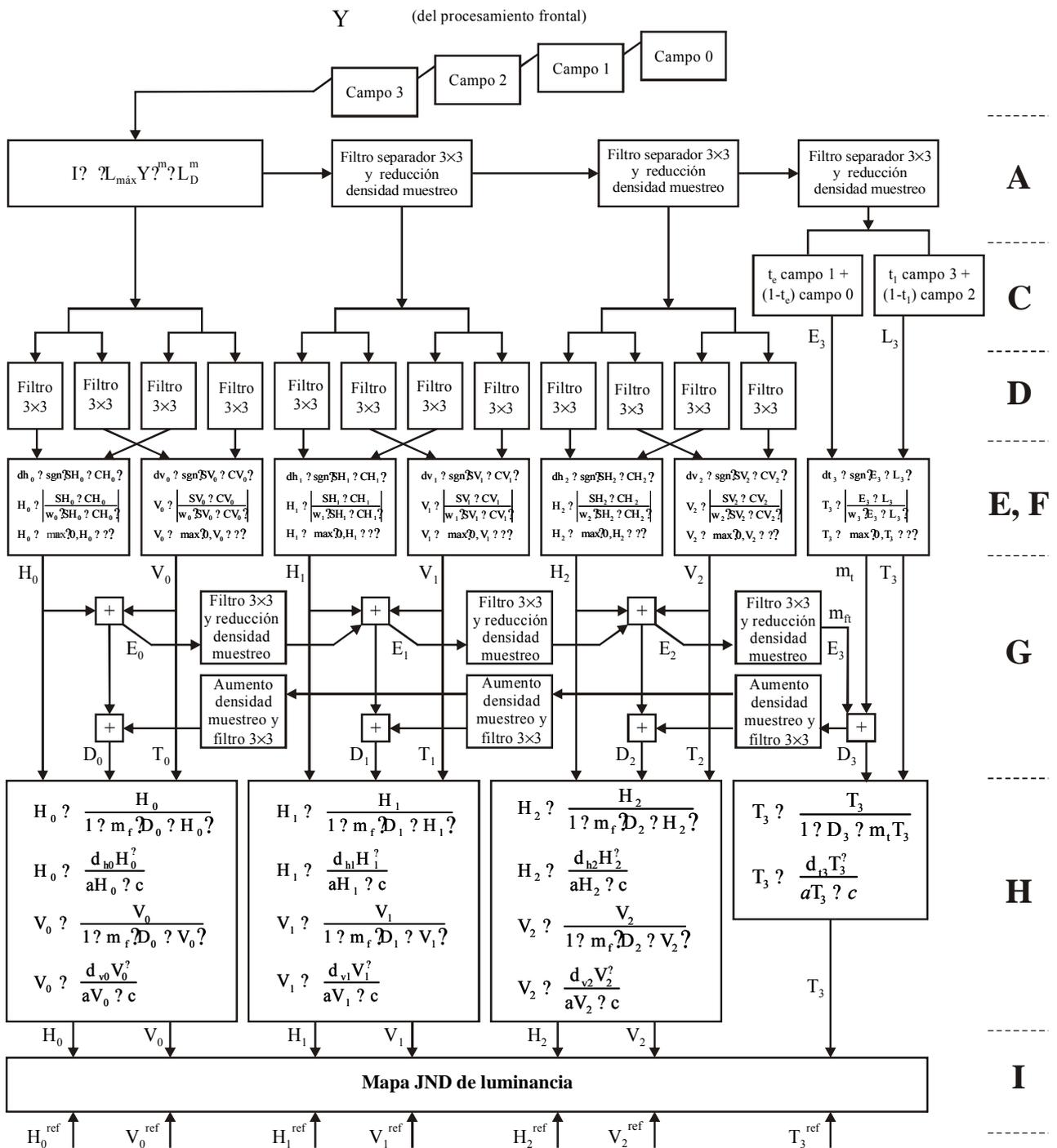
$$I_3(i,j) = [L_{\text{máx}} Y_3(i,j)]^m + L_d^m \quad (\text{II-14})$$

A  $L_{\text{máx}}$ , luminancia máxima de la pantalla, se le da el valor de  $100 \text{ cd/m}^2$ . Los valores de  $L_d$  y  $m$  se han escogido de modo que concuerden con los datos de detección del contraste para niveles de luminancia comprendidos entre 0,01 y  $100 \text{ ft-L}$ .

### II.5.2.2 Descomposición piramidal de la luminancia

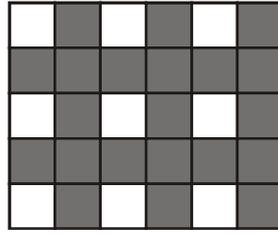
La descomposición espacial a cuatro niveles de resolución se efectúan mediante un método computacional de gran eficacia denominado procesamiento piramidal, que difumina y reduce la densidad de muestreo de la imagen en un factor de 2 a cada nivel sucesivo de resolución descendente.

<sup>23</sup> El canal de luminancia  $L^*$  de CIELUV no se utiliza en el procesamiento de la luminancia, sino que se sustituye por una no linealidad visual para la que se ha calibrado el modelo de visión en un rango de valores de luminancia. No obstante,  $L^*$  se utiliza en el procesamiento de la crominancia para crear una métrica de crominancia aproximadamente uniforme que sea conocida por los diseñadores de pantallas.



**Figura II.13 – Detalle del procesamiento de la luminancia**

A la imagen original de resolución completa se le denomina nivel cero de la pirámide,  $G_0 = I_3(i,j)$ . Los niveles subsiguientes, de menor resolución, se obtienen mediante una operación denominada REDUCE que funciona del siguiente modo. Un filtro paso bajo de tres derivaciones con pesos (1,2,1)/4 se aplica a  $G_0$  secuencialmente en cada dirección de la imagen para generar una imagen difuminada. A continuación se reduce la densidad de muestreo de la imagen en un factor de 2 (o sea, se suprime uno de cada dos píxeles como se muestra en la Figura II.14 con los píxeles a suprimir sombreados) para crear el siguiente nivel,  $G_1$ .



J.144.APPX.II\_FII.14

**Figura II.14 – Reducción de la densidad de muestreo de la imagen con supresión de los píxeles sombreados**

Si  $fds1()$  representa la operación de filtrado y reducción de la densidad de muestreo correspondiente a un nivel piramidal, el proceso REDUCE puede representarse por la siguiente expresión:

$$G_{i+1} = fds1(G_i), \text{ para } i=1, 2, 3. \quad (\text{II-15})$$

El proceso REDUCE se aplica reiteradamente cada nuevo nivel (como describen Burt y Adelson, 1983).

Del mismo modo, se define la operación contraria EXPAND que aumenta la densidad de muestreo y efectúa un filtrado con el mismo núcleo de  $3 \times 3$ . Esta operación se representa por  $usf1()$ , y se utiliza en II.5.2.5 y II.5.2.6.

Los núcleos de los filtros  $fds1$  y  $usf1$  en cada dirección (horizontal y vertical) son  $k_d [1,2,1]$  y  $k_u [1,2,1]$ , respectivamente, escogiéndose las constantes  $k_d$  y  $k_u$  de modo que se conserven valores de campo uniforme. Para  $fds1$  la constante  $k_d = 0,25$ , mientras que para  $usf1$  el valor de la constante es  $k_u = 0,5$  (debido a los ceros que se generan en el proceso de aumento de la densidad de muestreo de la imagen). Para implementar  $usf1$  como operación in situ, se sustituye el núcleo por la interpolación lineal equivalente a fin de sustituir los valores nulos. No obstante, para mayor simplicidad conceptual, seguiremos refiriéndonos a la operación como "filtrado de aumento de la densidad de muestreo".

### II.5.2.3 Filtrado espacial y temporal de la luminancia

Se aplican filtros espaciales orientados (centro y campo circundante) a las imágenes de los niveles 0, 1 y 2 para el campo 3. En el nivel de menor resolución (nivel 3), se combinan linealmente el primer par de campos y el último para formar imágenes adelantadas y atrasadas, respectivamente.

#### II.5.2.3.1 Filtrado espacial

Los filtros central y circundante son filtros de  $3 \times 3$  separables y proporcionan todas las combinaciones de orientación: centro vertical (CV, *center vertical*), centro horizontal (CH, *center horizontal*), circundante vertical (SV, *surround vertical*) y circundante horizontal (SH, *surround horizontal*). Los núcleos del filtro son los siguientes:

$$\text{CH} = \begin{bmatrix} 000 \\ 242 \\ 000 \end{bmatrix}; \quad \text{SH} = \begin{bmatrix} 121 \\ 000 \\ 121 \end{bmatrix}; \quad \text{CV} = \begin{bmatrix} 020 \\ 040 \\ 020 \end{bmatrix}; \quad \text{SV} = \begin{bmatrix} 101 \\ 202 \\ 101 \end{bmatrix} \quad (\text{II-16})$$

### II.5.2.3.2 Filtrado temporal

Las imágenes adelantada (e = early) y atrasada (l = late) del nivel 3 son, respectivamente, las siguientes

$$E_3 = t_e I_{3,1}(i, j) + (1 - t_e) I_{3,0}(i, j) \quad (\text{II-17})$$

$$L_3 = t_l I_{3,3}(i, j) + (1 - t_l) I_{3,2}(i, j) \quad (\text{II-18})$$

Las constantes  $t_e$  y  $t_l$  para 60 Hz son distintas de las correspondientes a 50 Hz.

### II.5.2.4 Cálculo del contraste de la luminancia

Las entradas son las imágenes central y circundantes  $CV_i$ ,  $CH_i$ ,  $SV_i$  y  $SH_i$  ( $i=0,1,2$  para los niveles piramidales 0, 1 y 2), para las imágenes adelantada y atrasada  $E_3$  y  $L_3$  (nivel 3) calculadas en II.5.2.3. La fórmula utilizada para calcular la relación de contraste es análoga a la del contraste Michelson,  $(L_{\text{máx}} - L_{\text{mín}})/(L_{\text{máx}} + L_{\text{mín}})$ , que ha demostrado ser satisfactoria para el modelado de la visión. Para las orientaciones horizontal y vertical, los respectivos contrastes, píxel a píxel, son

$$\frac{(SH_i - CH_i)}{w_i(CH_i + SH_i)} \text{ y } \frac{(SV_i - CV_i)}{w_i(CV_i + SV_i)} \quad (\text{II-19})$$

Análogamente, la relación de contraste para la componente temporal es

$$\frac{(E_3 - L_3)}{w_3(E_3 + L_3)} \quad (\text{II-20})$$

Los valores de  $w_i^{-1}$  para  $i = 0,1,2,3$  se determinaron por calibración.

Las imágenes de respuesta de contraste se calculan como versiones recortadas de los valores absolutos de las cantidades definidas por las dos ecuaciones anteriores. Estas cantidades se calculan del siguiente modo:

$$H_i = \text{máx} \left( 0, \left| \frac{(SH_i - CH_i)}{w_i(CH_i + SH_i)} \right| - \varepsilon \right), V_i = \text{máx} \left( 0, \left| \frac{(SV_i - CV_i)}{w_i(CV_i + SV_i)} \right| - \varepsilon \right) \quad (\text{II-21})$$

$i = 0,1,2$ , y

$$T_3 = \text{máx} \left( 0, \left| \frac{(E_3 - L_3)}{w_3(E_3 + L_3)} \right| - \varepsilon \right), \text{siendo } \varepsilon = 0,75. \quad (\text{II-22})$$

El signo algebraico de cada valor de píxel de la relación de contraste antes de la operación de valor absoluto (pasos E, F de la Figura II.13) debe conservarse para utilizarlo más adelante en el paso H.

### II.5.2.5 Enmascaramiento de contraste de la luminancia

El enmascaramiento de contraste es una función no lineal que se aplica a cada una de las respuestas de contraste calculadas en II.5.2.4. Modela el efecto de una estructura espacial y temporal en la secuencia de imágenes de referencia sobre la discriminación de la distorsión en la secuencia de imágenes de prueba.

Supongamos, por ejemplo, una imagen de prueba y otra de referencia, que difieren en una onda sinusoidal espacial de pequeña amplitud. Se sabe que esta diferencia es más visible cuando ambas imágenes tienen en común una onda sinusoidal de medio contraste y de la misma frecuencia espacial, que si ambas imágenes contienen un campo uniforme. No obstante, si el contraste de la onda sinusoidal común es demasiado grande, la diferencia entre las imágenes resulta menos aparente. Por otra parte, las ondas sinusoidales de otras frecuencias espaciales pueden repercutir en

la visibilidad de la diferencia de contraste. Este comportamiento puede modelarse mediante una no linealidad que sea una sigmoide para bajas energías de contraste y una función exponencial creciente para altas energías de contraste. Además, en la visión humana, pueden observarse aproximadamente las reglas siguientes. Cada canal se enmascara a sí mismo, las frecuencias espaciales altas enmascaran a las bajas (aunque no al revés), y el parpadeo temporal enmascara la sensibilidad de contraste espacial (y al revés).

Como respuesta a estas propiedades de la visión, se utiliza en el presente modelo la siguiente forma de no linealidad (aplicada píxel a píxel):

$$T(y, D_i) = \frac{d_y Z_i^\beta}{az_i + c} \quad (\text{II-23})$$

$$\text{siendo } z_i = \frac{y}{[1 + m_f(D_i - y)]} \text{ para } i = 0, 1, 2, \text{ y } z_3 = \frac{y}{(1 + D_3 - m_t y)}$$

Aquí,  $y$  es el contraste que ha de enmascarse: espacial,  $H_i$  o  $V_i$  (ecuación II-21) o temporal ( $T_3$ ) (ecuación II-22). La cantidad  $D_i$  se refiere (píxel a píxel) a una imagen que depende del nivel piramidal al que pertenece  $y$ . Las cantidades  $B$ ,  $a$ ,  $c$ ,  $m_f$  y  $m_t$  se determinaron por calibración.  $d_y$  es el signo algebraico del contraste  $y$  que se guardó antes de obtener el valor absoluto.

El cálculo de  $D_i$  requiere la construcción de la pirámide (filtrado seguido de reducción de la densidad de muestreo) y la reconstrucción de la pirámide (aumento de la densidad de muestreo seguido de filtrado). Esto puede verse en la Figura II.13 y en las ecuaciones siguientes. En estas ecuaciones,  $fds1()$  representa el filtrado de  $3 \times 3$  seguido de la reducción de densidad de muestreo en un nivel piramidal, mientras que  $usf1()$  representa el aumento de la densidad de muestreo en un nivel piramidal seguido del filtrado de  $3 \times 3$  (véase el final del II.5.2.2). En primer lugar se calcula la matriz  $E_0$  como

$$E_0 = H_0 + V_0 \quad (\text{II-24})$$

A continuación, se calculan reiteradamente las matrices  $E_i$  para  $i = 1, 2$  del siguiente modo:

$$E_i = H_i + V_i + fds1(E_{i-1}), \text{ para } i = 1, 2 \quad (\text{II-25})$$

$$E_3 = fds1(E_2) \quad (\text{II-26})$$

A continuación las matrices  $E_i$  se combinan con la imagen temporal de contraste  $T_3$  y con las imágenes  $T_i$  para generar las matrices denominadoras de contraste  $D_i$ , del siguiente modo:

$$D_3 = m_t T_3 + m_{ft} fds1(E_2), \quad (\text{II-27})$$

$$T_2 = usf1(D_3), \quad T_i = usf1(T_{i+1}), \text{ para } i = 1, 0, \text{ y}$$

$$D_i = E_i + T_i, \text{ para } i = 0, 1, 2 \quad (\text{II-28})$$

Aquí, el parámetro  $m_{ft}$  modula la intensidad con la que se enmascara el canal de luminancia temporal (parpadeo) por todos los canales espaciales de luminancia juntos; mientras que el parámetro  $m_t$  modula la intensidad con la que se enmascara el canal de luminancia espacial por el canal de luminancia temporal (parpadeo).

El procesamiento anterior demuestra que las frecuencias espaciales superiores enmascaran a las inferiores (dado que  $D_i$  está influenciada por los niveles piramidales menores o iguales que  $i$ ), y que el canal temporal enmascara y es enmascarado por todos los canales espaciales. Esto corresponde aproximadamente a la observación psicofísica. Como se podrá observar, las cantidades  $D_i$ ,  $i = 0, 1, 2$ , también enmascaran los contrastes de crominancia (aunque no al revés).

### II.5.2.6 Construcción del mapa JND de luminancia

La construcción descrita a continuación se aplica a todas las imágenes de contraste enmascarado generadas en el paso H anterior (véase la Figura II.13).

- las imágenes de las pirámides H y V (o sea, las imágenes  $H_0, V_0, H_1, V_1, H_2$  y  $V_2$ );
- la imagen  $T_3$  (la de resolución a nivel 3);
- las correspondientes imágenes obtenidas de la secuencia de referencia (afectadas del superíndice<sup>ref</sup> en la Figura II.13).

Los cuatro primeros pasos del proceso siguiente se aplican a las citadas imágenes por separado. En su exposición, se representa por  $X$  cualquiera de las imágenes obtenidas de la secuencia de prueba y por  $X^{\text{ref}}$  la imagen correspondiente obtenida de la secuencia de referencia. De acuerdo con esta notación, los pasos resultantes son los siguientes:

- Separar la imagen  $X$  en dos imágenes rectificadas en media onda, una para los contrastes positivos y otra para los negativos. En la imagen de contrastes positivos (denominada  $X_+$ ), los signos del contraste  $X$  (almacenados por separado en la etapa E) se utilizan para asignar ceros a todos los píxeles de  $X_+$  que tengan contraste negativo. Lo contrario ocurre en la imagen de contraste negativo  $X_-$ .
- Para cada imagen  $X_+$  y  $X_-$ , ejecutar la operación de agrupamiento local sugerida por la psicofísica efectuando la convolución de la imagen con el núcleo  $0,25(1,2,1)$  tanto horizontal como verticalmente.
- Reducir la densidad de muestreo de las imágenes resultantes en un factor de 2 en cada dirección, y suprimir la redundancia resultante del agrupamiento del paso anterior. Suponiendo que el procesamiento aplicado a  $X$  se haya efectuado también para la imagen de referencia correspondiente  $X^{\text{ref}}$ , calcular píxel a píxel las imágenes de diferencias absolutas  $|X_+ - X_+^{\text{ref}}|$  y  $|X_- - X_-^{\text{ref}}|$ . Las imágenes resultantes forman los mapas JND.

Una vez completado este proceso para todos los pares  $X, X^{\text{ref}}$ , deberán repetirse los aumentos de densidad de muestreo, filtrados y añadirse todas las imágenes al nivel necesario para calcular las mediciones resumidas. Esto se lleva a cabo del siguiente modo:

- inicializar una imagen de sumas acumulativas que contenga la suma de la  $Q$ -ésima potencia de las imágenes del nivel 3 obtenidas de  $T_3, T_3^{\text{ref}}$ :  $|T_3 + - T_3^{\text{ref}}|^Q$  y  $|T_3 - T_3^{\text{ref}}|^Q$ , siendo  $Q$  igual a 2;
- aumentar la densidad de muestreo/filtrar la imagen de sumas acumulativas para componer una imagen de nivel 2;
- actualizar la imagen de suma acumulativa añadiéndole la  $Q$ -ésima potencia de las imágenes de nivel 2 obtenidas de  $H_2, H_2^{\text{ref}}, V_2$  y  $V_2^{\text{ref}}$ ;
- aumentar la densidad de muestreo/filtrar la imagen de suma acumulativa para componer una imagen de nivel 1;
- actualizar la imagen de suma acumulativa añadiéndole la  $Q$ -ésima potencia de las imágenes de nivel 1 obtenidas de  $H_1, H_1^{\text{ref}}, V_1$  y  $V_1^{\text{ref}}$ ;
- aumentar la densidad de muestreo/filtrar la imagen de suma acumulativa para componer una imagen de nivel 0; y
- actualizar la imagen de suma acumulativa añadiéndole la  $Q$ -ésima potencia de las imágenes de nivel 0 obtenidas de  $H_0, H_0^{\text{ref}}, V_0$  y  $V_0^{\text{ref}}$ . Enviar esta imagen directamente al procesamiento resumen (véase la Figura II.9 y II.5.4).

Obsérvese que tras este proceso, la imagen resultante tiene la mitad de resolución que la original. De modo semejante, obsérvese que cada índice de nivel de pirámide de esta cláusula se refiere al nivel piramidal del que se obtuvo originalmente, que tiene dos veces la resolución de la asociada a dicho nivel tras el filtrado/reducción de la densidad de muestreo.

## II.5.2.7 Procesamiento de luminancia de media altura

Si las imágenes de media altura deben pasarse directamente a imágenes de altura real sin relleno de ceros, debe modificarse el anterior procesamiento de la luminancia para tener en cuenta que la resolución vertical intrínseca sólo es la mitad de la resolución horizontal intrínseca. En la Figura II.15 se resume el algoritmo de la luminancia para media altura.

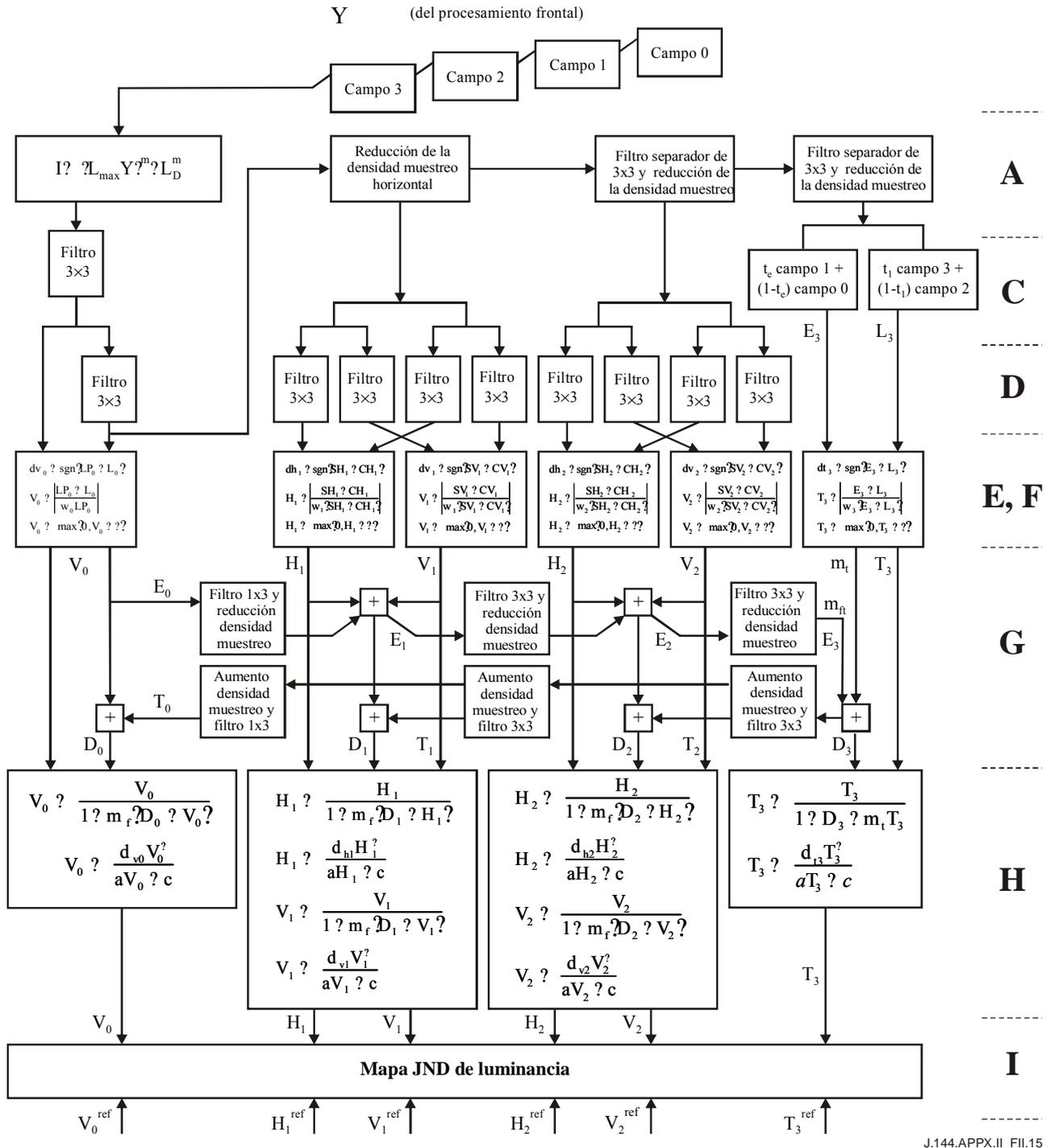


Figura II.15 – Detalle del procesamiento de la luminancia (media altura)

La comparación entre este diagrama y el correspondiente de altura completa (Figura II.13) pone de manifiesto las siguientes diferencias básicas:

- 1) Desaparece el canal horizontal de máxima resolución,  $H_0$ .
- 2) Tras el paso A, la imagen de la máxima resolución se somete a un filtrado paso bajo verticalmente (es decir, en las columnas) con un filtro "Kell" de  $3 \times 1$  con pesos  $(1/8, 3/4, 1/8)$ . Esta operación corresponde al filtrado conjunto del filtro de desentrelazado supuesto, y del realizado por las componentes verticales de los filtros de  $3 \times 3$  en el paso D del algoritmo de altura completa. La imagen resultante filtrada verticalmente,  $L_0$ , se filtra a continuación horizontalmente con un filtro de  $1 \times 3$  (núcleo  $0,25[1,2,1]$ ). La imagen resultante,  $LP_0$ , es una versión sometida a un filtrado paso bajo horizontal, de  $L_0$ .
- 3)  $L_0$  y  $LP_0$  se combinan en los pasos E y F para generar una respuesta orientada a paso banda  $(LP_0 - L_0)$  dividida por paso bajo  $(LP_0)$ , análoga a las respuestas  $(S - C)/(S + C)$  de los otros canales orientados.
- 4) Se reduce la densidad de muestreo horizontal de la imagen  $LP_0$  (imagen de media altura de  $720 \times 240$  píxeles) en el paso A, para convertirla en una imagen de media resolución ( $360 \times 240$ ) y altura completa. El procesamiento de esta imagen y el de los tres niveles piramidales restantes continúa como en las opciones de altura completa.
- 5) En el paso G, el aumento y reducción de densidad de muestreo entre las imágenes de media altura del nivel 0 y las imágenes de altura completa del nivel 1 se efectúa por filtrado de  $1 \times 3$ /reducción de densidad de muestreo horizontal (que se indica como filtro  $1 \times 3$  y reducción densidad muestreo) y aumento de la densidad de muestreo horizontal/filtrado de  $1 \times 3$  respectivamente. La reducción de muestreo horizontal supone la división por dos de la dimensión horizontal; es decir, la supresión de una columna de la imagen de cada dos. El aumento de densidad de muestreo horizontal significa la introducción de una columna de ceros entre cada dos de la imagen existente. El núcleo del filtro tras el aumento de densidad de muestreo se define como  $0,5 [1,2,1]$ , por los motivos enunciados al final del II.5.2.2.

Por otra parte, en la construcción del Mapa JND, el filtrado de  $3 \times 3$  y la reducción de densidad de muestreo de  $V_0$  se sustituye por un filtrado de  $1 \times 3$  y una reducción de la densidad de muestreo horizontal.

### II.5.3 Procesamiento de la crominancia

Las cláusulas II.5.3.1 a II.5.3.6 describen el procesamiento de altura completa. La cláusula II.5.3.7 trata de las modificaciones necesarias para el procesamiento de media altura.

#### II.5.3.1 Descomposición de la pirámide de crominancia

Además de la pirámide con niveles 0, 1, 2 (para el cálculo de  $Y$  en el procesamiento de luminancia), hay que utilizar pirámides con los niveles 3, 4, 5 y 6 para el cálculo de  $u^*$  y  $v^*$ . Así pues:

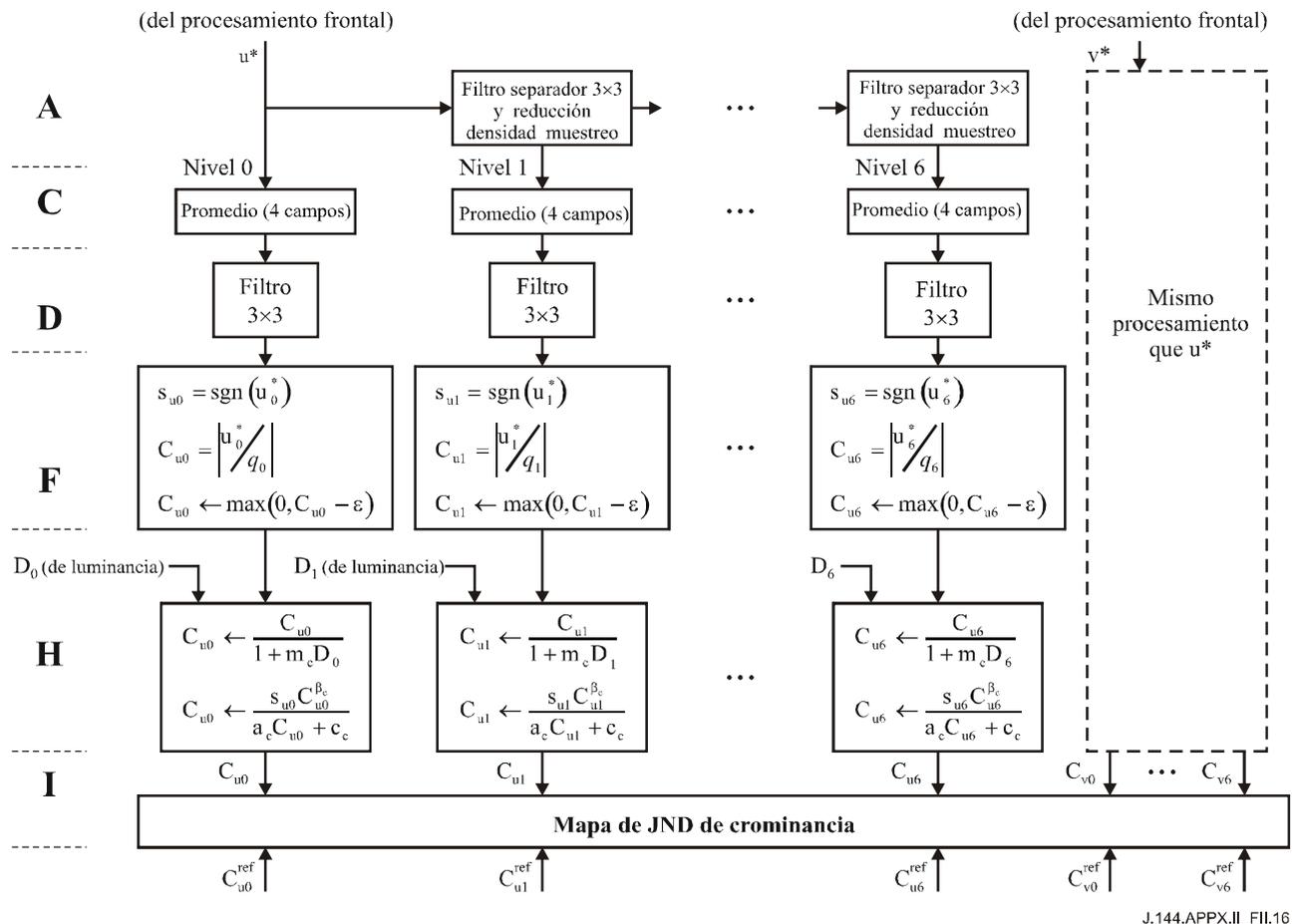
$$u_0 = u^*, v_0 = v^*, u_i = fds1(u_{i-1}), v_i = fds1(v_{i-1}), i = 1, \dots, 6, \quad (\text{II-29})$$

donde  $fds1()$  representa la operación de filtrado y reducción de densidad de muestreo definida en II.5.2.2. Véase la Figura II.16.

La resolución espacial del canal de crominancia de máxima resolución (nivel 0) se escoge igual a la del canal de luminancia del nivel 0 debido a que la resolución depende de la separación entre píxeles, y no de la separación entre receptores. La separación entre receptores es de 0,007 grados del ángulo visual, mientras que la separación entre píxeles es de tan sólo 0,03 -obtenida en una pantalla con 480 píxeles de altura, y una distancia de observación de cuatro veces su altura. Asimismo, la resolución del canal cromático azul-amarillo viene limitada por el hecho de que el sistema visual es tritanope (ciego al azul) para luces que se extienden menos de  $2'$  (o sea, 0,033 grados) aproximadamente del ángulo visual. La resolución de píxel de 0,03 grados del ángulo

visual es tan próxima al mayor de estos valores que es conveniente hacer iguales las resoluciones de los canales de luminancia y crominancia.

La pirámide de crominancia llega hasta el nivel 6 en vez de hasta el 2. Esto confirma el hecho de que los observadores perciben diferencias entre campos de color grandes espacialmente uniformes.



NOTA –  $D_3, \dots, D_6$  se calcula por sucesivos filtrados y reducciones de la densidad de muestreo de  $D_2$  (a partir de la luminancia).

**Figura II.16 – Detalle del procesamiento de la crominancia<sup>24</sup>**

### II.5.3.2 Procesamiento temporal de la crominancia

Para cada nivel de resolución  $i$ , calcular el promedio de cuatro campos de las imágenes  $u_i$ , y asimismo de las imágenes  $v_i$ , con pesos de derivaciones (0,25, 0,25, 0,25, 0,25), es decir:

$$u_i \leftarrow \frac{1}{4} \sum_{j=0}^3 u_i^j \quad v_i \leftarrow \frac{1}{4} \sum_{j=0}^3 v_i^j \quad (\text{II-30})$$

siendo  $j$  el índice del campo.

Este paso refleja el filtrado temporal de paso bajo intrínseco de los canales de color, y sustituye al procesamiento adelantado-atrasado del canal de luminancia temporal.

<sup>24</sup> Las etiquetas de los pasos son las especificadas para mantener la coherencia las de los pasos del procesamiento de la luminancia.

### II.5.3.3 Filtrado espacial de la crominancia

Aplicar un filtro espacial no orientado de Laplace a cada una de las imágenes  $u_i$  y  $v_i$ . El filtro utilizado en cada caso tiene el siguiente núcleo de  $3 \times 3$ :

$$1/4 \begin{bmatrix} 1 & 2 & 1 \\ 2 & -12 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad (\text{II-31})$$

Este núcleo se ha escogido de modo que el peso total sea cero y responda con una intensidad máxima de 1 a cualquier borde recto entre dos zonas uniformes cuya diferencia sea la unidad. (La respuesta máxima se alcanza en un borde horizontal o vertical.) Esto convierte las imágenes  $u_i$  y  $v_i$  en mapas de diferencias de crominancia evaluadas en unidades (JND) de espacio de color uniforme.

### II.5.3.4 Cálculo del contraste de crominancia

Adoptar directamente las imágenes  $u_i$  y  $v_i$  del paso D como pirámides de contraste de crominancia, e interpretarlas de modo similar a los contrastes Michelson calculados en el paso E del modelo de luminancia. Al igual que los contrastes de luminancia, los contrastes de croma se calculan mediante comparaciones intraimagen afectadas por pirámides de Laplace. Así como la diferencia de Laplace dividida por una media espacial representa el contraste de Michelson, que mediante la ley de Weber supone un valor constante en el nivel 1-JND (umbral de detección), la pirámide de Laplace operando sobre  $u_i$  y  $v_i$  tiene una interpretación 1-JND. Como ocurría con el modelo de luminancia, esta interpretación debe modificarse en el curso de la calibración para recoger la interacción de todas las partes del modelo y reflejar el hecho de que los estímulos que provocan la respuesta 1-JND no sean sencillos en términos del modelo.

A continuación se dividen, nivel a nivel, las imágenes de la pirámide de contrastes por siete constantes  $q_i$  ( $i=0, \dots, 6$ ) cuyos valores se determinan por calibración. Estas constantes son análogas a las cantidades  $w_i$  ( $i=0,1,2,3$ ) del modelo de luminancia.

Calcular los valores absolutos recortados de todos los contrastes  $u_i$  y  $v^*$  [donde  $\text{clip}(x) = \max(0, x - \epsilon)$ ], siendo  $\epsilon = 0,75$ . Consérvense los signos algebraicos hasta el paso H en el que se volverán a incorporar dichos signos. Esto evita la posibilidad de registrar 0 JND entre dos imágenes distintas debido a la ambigüedad provocada por la pérdida del signo en la operación de valor absoluto. Los resultados son dos pirámides de contraste de crominancia  $C_u, C_v$ .

### II.5.3.5 Enmascaramiento de contraste de la crominancia

Adoptar los niveles piramidales del denominador  $D_m$  ( $m=0, 1, 2$ ) directamente del paso G del modelo de luminancia, sin ninguna modificación. Para los niveles 3, ..., 6 ejecutar un filtrado y una reducción de densidad de muestreo secuenciales de  $D_2$  con arreglo al mismo método que en el procesamiento de luminancia, pero sin añadir nuevos términos. Estos valores  $D_m$  se utilizan en el paso H en consideración a la teoría de la perturbación. Debido a que cabe esperar que los efectos de la luminancia predominen sobre los de la crominancia en la mayor parte de los casos, el modelo de crominancia puede considerarse como una perturbación de primer orden del modelo de luminancia. Por consiguiente, los efectos de la luminancia (la  $D_m$ ) se pueden modelar como crominancia de enmascaramiento, aunque no al revés.

Utilizar la pirámide denominadora del canal de luminancia  $D_m$  y la misma forma funcional que se utiliza para que el transductor de luminancia enmascare las pirámides de contraste de crominancia, para todos los niveles piramidales  $m=0, \dots, 6$ :

$$C_{um} \leftarrow \frac{s_{um} z_{um}^{\beta_c}}{a_c C_{um} + c_c} \quad (\text{II-32})$$

$$\text{siendo } z_{um} = \frac{C_{um}}{(1 + m_c D_i)}$$

y  $D_i$  una versión filtrada de densidad de muestreo reducida de  $D_2$  cuando  $i > 2$ . Análogamente,

$$C_{vm} \leftarrow \frac{s_{vm} z_{vm}^{\beta_c}}{a_c C_{vm} + c_c} \quad (\text{II-33})$$

$$\text{siendo } z_{vm} = \frac{C_{vm}}{(1 + m_c D_i)}$$

Obsérvese que el signo algebraico suprimido en el paso F se ha reincorporado mediante los factores  $s_{um}$  y  $s_{vm}$ . Esto genera pirámides de contraste enmascaradas para  $u_i$  y  $v_i$ . La calibración determina los valores  $a_c$ ,  $c_c$ ,  $\beta_c$ ,  $m_c$  y  $m_f$ .

### II.5.3.6 Construcción del mapa JND de crominancia

La construcción del mapa JND de crominancia obedece a un proceso totalmente análogo al de la construcción del mapa JND de luminancia (véase II.5.2.6). En este caso, el procedimiento se aplica a todas las imágenes de crominancia de contraste enmascarado generadas en el paso H anterior (véase la Figura II.16).

- las imágenes  $C_{u0}$ ,  $C_{v0}$ , ...,  $C_{u6}$ ,  $C_{v6}$
- las imágenes correspondientes derivadas de la secuencia de referencia (representadas con el superíndice <sup>ref</sup> en la Figura II.11).

Los tres primeros pasos del siguiente proceso se aplican a las anteriores imágenes por separado. Al tratar de ellas, representaremos por  $X$  cualquiera de las imágenes obtenidas de la secuencia de prueba, y por  $X^{\text{ref}}$  la imagen correspondiente obtenida de la secuencia de referencia. De acuerdo con esta notación, los pasos resultantes son los siguientes:

- Separar la imagen  $X$  en dos imágenes rectificadas de media onda, una para los contrastes positivos y otra para los negativos. En la imagen de contrastes positivos (denominada  $X_+$ ), los signos del contraste  $X$  (almacenados por separado en la etapa E) se utilizan para asignar ceros a todos los píxeles de  $X_+$  que tengan contrastes negativos. Lo contrario sucede en la imagen de contrastes negativos  $X_-$ .
- Para cada imagen  $X_+$  y  $X_-$  ejecutar la operación de agrupamiento local sugerida por la psicofísica mediante convolución de la imagen con el núcleo  $0,5(1,2,1)$ , tanto horizontal como verticalmente. A continuación, reducir la densidad de muestreo de las imágenes resultantes en un factor de 2 en cada dirección, para eliminar la redundancia resultante del agrupamiento.
- Suponiendo que el mismo procesamiento efectuado sobre  $X$  se haya realizado para la correspondiente imagen de referencia  $X^{\text{ref}}$ , calcular píxel a píxel las imágenes de diferencia absolutas  $|X_+ - X_+^{\text{ref}}|$  y  $|X_- - X_-^{\text{ref}}|$ . Las imágenes resultantes son los mapas JND.

Una vez completado este proceso para todos los pares  $X$ ,  $X^{\text{ref}}$ , efectuar repetidamente el aumento de densidad de muestreo, filtrado y adición o maximización de todas las imágenes en el nivel necesario para calcular las mediciones resumen. Esto se realiza del siguiente modo:

- Inicializar una imagen de sumas acumulativas que contenga la suma de las  $Q$ -ésimas potencias de las imágenes de nivel 6 obtenidas de  $C_{u6}$ ,  $C_{u6}^{\text{ref}}$ ,  $C_{v6}$ , y  $C_{v6}^{\text{ref}}$ . Aquí,  $Q = 2$ .

A continuación ejecutar los dos pasos siguientes desde el nivel piramidal  $m = 5$  hasta el 0 en pasos descendentes de 1.

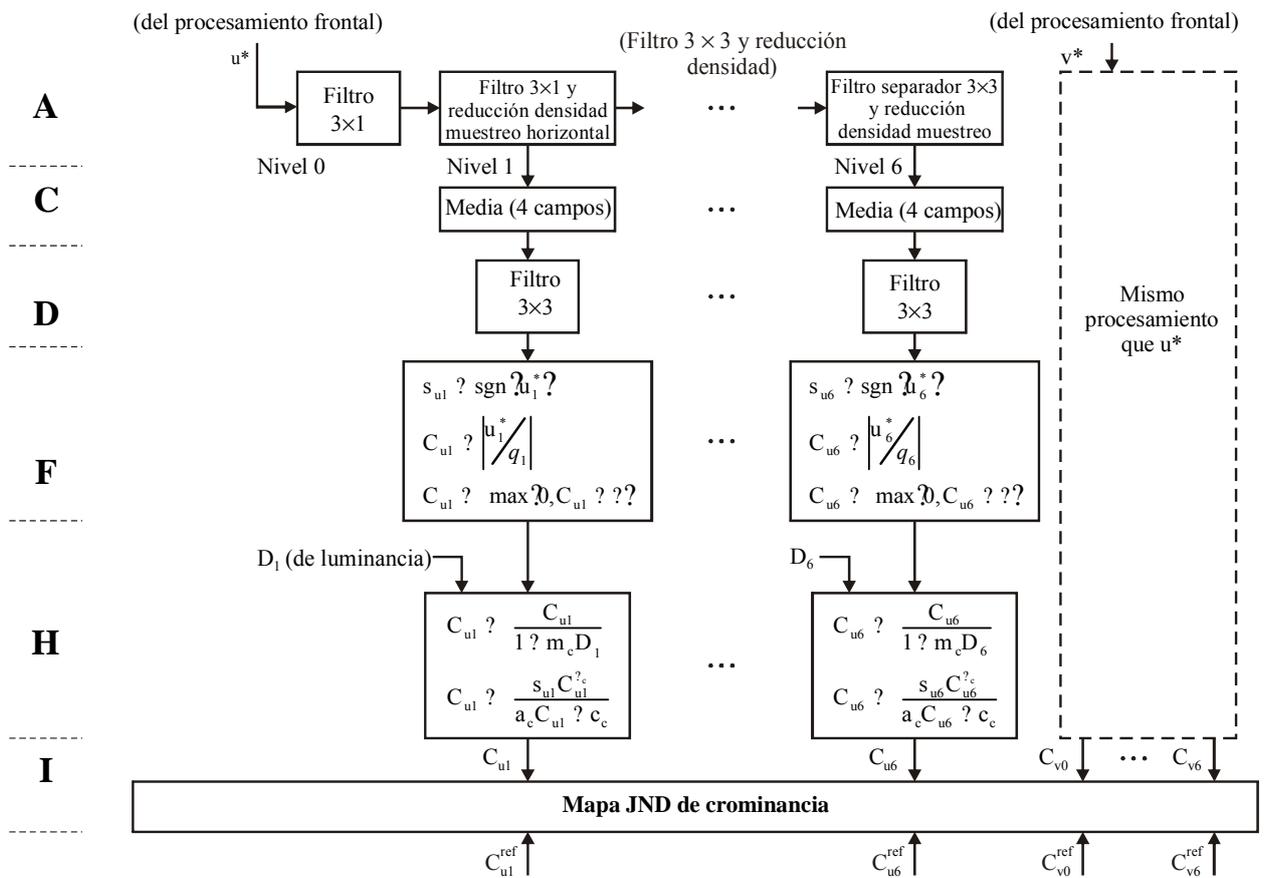
- Aumentar la densidad de muestreo/filtrar la imagen de suma acumulativa para componer una imagen de nivel  $m$ .

- Actualizar la imagen de suma acumulativa añadiéndole las  $Q$ -ésimas potencias de las imágenes del nivel  $n$  obtenidas de  $C_{um}$ ,  $C_{um}^{ref}$ ,  $C_{vm}$  y  $C_{vm}^{ref}$ .

Al igual que en el procesamiento de la luminancia, una vez efectuadas estas operaciones, la imagen resultante tiene la mitad de resolución que la original. Obsérvese que en esta cláusula cada índice de nivel piramidal se refiere al nivel piramidal del que se obtuvo originalmente, que es dos veces la resolución asociada a dicho nivel tras el filtrado/reducción de densidad de muestreo. La imagen de nivel 0 se envía directamente al procesamiento resumen (véase la Figura II.9 y II.5.4).

### II.5.3.7 Procesamiento de la crominancia de media altura

Si hay que pasar las imágenes de media altura a imágenes de altura real directamente sin relleno de ceros, el anterior procesamiento de crominancia debe modificarse para tener en cuenta que la resolución vertical intrínseca sólo es la mitad de la resolución horizontal intrínseca. La Figura II.17 resume el procesamiento de crominancia para el algoritmo de media altura.



J.144.APPX.II.FII.17

NOTA –  $D_3, \dots, D_6$  se calculan por filtrado y reducción de la densidad de muestreo sucesivos de  $D_2$  (de luminancia).

**Figura II.17 – Detalle del procesamiento de la crominancia (media altura)**

La comparación entre este diagrama y el de entrelazado de altura completa (Figura II.11) pone de manifiesto las siguientes diferencias básicas:

- 1) Desaparecen los canales de crominancia de máxima resolución,  $u_0^*$  y  $v_0^*$ . Como la sensibilidad de crominancia es baja a frecuencias espaciales altas, no se pierde nada importante en este paso.

- 2) En el Paso A, para generar las imágenes de crominancia de la siguiente resolución más alta  $u_1^*$  y  $v_1^*$ , se aplica verticalmente (es decir a las columnas) un filtro "Kell" de paso bajo cuyo núcleo tiene los pesos (1/8, 3/4, 1/8). Esta operación corresponde al filtrado conjunto del filtro de desentrelazado supuesto y del efectuado por las componentes verticales de los filtros de  $3 \times 3$  en el Paso D del algoritmo de altura completa. Las imágenes resultantes filtradas verticalmente se filtran a continuación horizontalmente con un núcleo 0,25(1,2,1) de un filtro de  $1 \times 3$ . Este filtrado de las imágenes  $u^*$  y  $v^*$  hace que la resolución de las imágenes de media altura sea isótropa. La resolución es la misma que la del nivel piramidal de altura completa 1.
- 3) Debido a que el tren de norma Q está totalmente acumulado en el nivel piramidal 1 del modelo de crominancia, el mapa JND de crominancia para mediciones resumidas sólo tiene la mitad del tamaño (tanto horizontal como verticalmente) del mapa de luminancia totalmente acumulada. Antes de combinar los mapas de crominancia y luminancia para producir el mapa JND total, debe llevarse el mapa de crominancia a la misma resolución que el mapa de luminancia. Para alcanzar este objetivo, se efectúa un aumento de la densidad de muestreo seguida de un filtrado de  $3 \times 3$  para producir el mapa JND de crominancia de mediciones resumidas.

#### II.5.4 Resúmenes de salida

Como se ha expuesto en anteriores cláusulas, los mapas JND de luminancia y crominancia que se pasan a la etapa resumen de salida son imágenes JND, y se representan con la mitad de la resolución de la imagen original. Esto aprovecha la redundancia propia de tener que ejecutar el agrupamiento en cada etapa de contraste enmascarado.

A continuación, los mapas JND de luminancia y crominancia,  $JND_L$  y  $JND_C$ , se combinan en un mapa JND de campo total,  $JND_T$ . La regla de combinación es una norma Q de Minkowski ( $Q = 2$ ), de forma semejante a la combinación de canales para producir los mapas  $JND_L$  y  $JND_C$ :

$$JND_T(i,j) = [JND_L(i,j)^Q + JND_C(i,j)^Q]^{1/Q} \quad (II-34)$$

A continuación, cada uno de los tres mapas JND (de luminancia, de crominancia, y de luminancia y crominancia combinadas) se reduce a un resumen de un sólo número, denominado valor índice de calidad de imagen (PQR). Los resúmenes de un sólo número se calculan mediante la norma Q de Minkowski como se indica a continuación.

Cada una de las imágenes JND de media resolución (tres para cada campo luminancia, crominancia y campo total) se reduce a una única medida de calidad de funcionamiento denominada PQR, mediante las fórmulas siguientes:

$$\begin{aligned}
 PQR_{luminancia} &= \left[ \left( \frac{1}{N_p} \right) \sum_{i,j} JND_L(i,j)^Q \right]^{1/Q} \\
 PQR_{crominancia} &= \left[ \left( \frac{1}{N_p} \right) \sum_{i,j} JND_C(i,j)^Q \right]^{1/Q} \\
 PQR_{total} &= \left[ \left( \frac{1}{N_p} \right) \sum_{i,j} JND_T(i,j)^Q \right]^{1/Q}
 \end{aligned} \quad (II-35)$$

donde la suma se efectúa para todos los píxeles del mapa JND,  $Q = 4$ , y  $N_p$  es el número de píxeles del mapa. De este modo se calculan tres mediciones resumidas correspondientes, respectivamente, a  $JND_L$ ,  $JND_C$  y  $JND_T$ , para cada campo  $k$  de una secuencia de vídeo.

A partir de los  $N$  valores  $PQR_{campo}^{25}$  de campo único de una secuencia de vídeo, se calcula una única medición de la calidad de funcionamiento  $PQR_N$  mediante la siguiente norma  $Q$  de Minkowski:

$$PQR_N = \left[ \left( \frac{1}{N} \right) \sum_k PQR_{campo}(k)^Q \right]^{1/Q} \quad (II-36)$$

NOTA – Los datos de índice subjetivo llevan ruido y no son fiables para secuencias de vídeo cortas (menos de 1/2 segundo o 15 cuadros). La correlación de las estimaciones PQR con los índices subjetivos para secuencias cortas es bastante débil.

## II.5.5 Procesamiento de los bordes de la imagen

Para minimizar los recortes y por lo tanto, para evitar artefactos en el marco, el método PQR sustituye el marco de la pantalla por un bisel gris de amplitud infinita, que no aumenta el tamaño real de la imagen en más de seis píxeles a cada lado de la misma. Este "bisel virtual" elimina la necesidad de recortar el mapa JND para evitar los artefactos del marco. El bisel gris infinito modela las condiciones de observación y por tanto puede considerarse que no genera artefactos. De acuerdo con esta interpretación, todo el mapa JND queda libre de contaminación por artefactos.

En esta cláusula se describe el algoritmo del marco. En la siguiente exposición, la imagen rellena con seis píxeles a cada lado se denomina *imagen con orla*, y la imagen sin relleno o su lugar dentro de la imagen con orla se denomina *imagen propiamente dicha*.

### II.5.5.1 Color del bisel

Dado que las operaciones de imagen son locales, el bisel virtualmente infinito puede implementarse eficazmente. A una distancia suficiente de la *imagen propiamente dicha*, el bisel infinito consiste en un conjunto de valores constantes e idénticos en cualquier etapa del modelo dada. El efecto de las operaciones de imagen, por ejemplo filtrado, sobre esta región constante puede calcularse a priori. Así pues, un marco estrecho (seis píxeles en la actual implementación) puede facilitar la adecuada transición de la *imagen propiamente dicha* al bisel infinito.

A la entrada, se le otorgan al bisel los valores  $Y' = 90$ ,  $U' = V' = 0$ . (El valor  $Y' = 90$  corresponde a la mitad del valor de fondo de la Rec. UIT-R BT.500-11 de 15% de la luminancia de pantalla máxima.) No obstante, el bisel no es necesario hasta después del procesamiento frontal, debido a que las interacciones espaciales que sobrepasan el marco de la imagen no tienen lugar hasta después de esta etapa. En el canal de luminancia no se añaden marcos (y por consiguiente tampoco valores de bisel) a las imágenes hasta después de la compresión de luminancia. En el canal de crominancia, los marcos se añaden tras el procesamiento frontal.

En el canal de luminancia, el primer valor de bisel tras la compresión de luminancia es:

$$primer\_bisel\_luminancia = \left[ L_{\max} \left( \frac{90}{255} \right)^{\gamma} \right]^m + L_d^m \quad (II-37)$$

En los canales  $u^*$  y  $v^*$  los valores del primer bisel son ambos 0.

<sup>25</sup> Aquí, *campo* representa un valor PQR de cualquiera de las tres secuencias: luminancia, crominancia o total.

Estos valores se propagan por las subsiguientes etapas del modelo de tres maneras:

- 1) Hay funciones que operan píxel a píxel sobre valores de bisel antiguos para producir valores de bisel nuevos. Por ejemplo, el valor de bisel resultante de la función potencia (ecuación II-23) es:

$$\text{bisel\_salida} = (\text{bisel\_entrada})^R \quad (\text{II-38})$$

- 2) Hay filtros espaciales de  $3 \times 3$  cuyas filas y columnas suman P y dan al valor del bisel de salida el de la entrada multiplicado por P.
- 3) Hay numeradores de función de contraste y filtros temporales de cuatro campos (cuyas derivaciones suman cero) y dan al bisel de salida el valor 0.

En la etapa de contraste y en las siguientes, se da al bisel el valor 0 en los canales de luminancia y crominancia – consecuencia natural de operar con un núcleo lineal de suma 0 en una matriz espacialmente constante.

Las tres categorías anteriores introducen algunas de las complejidades necesarias para entender e implementar el algoritmo de marco. En la cláusula siguiente, se introduce un nuevo nivel de detalle.

### II.5.5.2 Integración de la imagen y el bisel

En las primeras etapas de la pirámide del modelo, no es necesario suministrar marco. La primera operación de marco en una imagen de entrada N por M consiste en rellenar la imagen con seis píxeles (a cada lado) con el valor de bisel adecuado (primer\_bisel\_luminancia para la imagen de luminancia comprimida, y 0 para las imágenes  $u^*$  y  $v^*$ ). La imagen con orla tiene las dimensiones  $(N + 12) \times (M + 12)$ . Para el k-ésimo nivel piramidal (donde k puede variar entre 0 y 7)<sup>26</sup>, la imagen con orla tiene las dimensiones  $(\lceil N/2^k \rceil + 12) \times (\lceil M/2^k \rceil + 12)$ , donde "[x]" representa el mayor entero de x.

Las imágenes en todos los niveles piramidales se registran recíprocamente en el ángulo superior izquierdo de la *imagen propiamente dicha*. Los índices de la *imagen propiamente dicha* tienen el siguiente recorrido  $0 = y = \text{altura}$ ,  $0 = x = \text{anchura}$ . El ángulo superior izquierdo de la *imagen propiamente dicha* siempre tiene los índices (0,0). Los índices de los píxeles del bisel toman valores de altura y anchura menores que 0. Por ejemplo, el píxel del bisel superior izquierdo es (-6, -6). En la dimensión x comenzando por el borde izquierdo de una imagen de anchura w (anchura de la imagen más el bisel  $w + 12$ ), los píxeles del bisel se indexan según  $x = (-6, -5, \dots, -1)$  y la imagen real se indexa como  $(0, 1, \dots, w-1)$  mientras que el índice del bisel derecho abarca  $(w, w + 1, \dots, w + 5)$ .

En una imagen con orla pueden presentarse cuatro casos en función de la etapa siguiente de procesamiento. En la descripción de las operaciones siguientes, se utilizan las líneas de una sola imagen para resumir el procesamiento espacial (entendiendo que sucede lo mismo en dirección vertical).

- a) *En las operaciones píxel a píxel.* Cuando la siguiente operación deba efectuarse píxel a píxel (por ejemplo, con una no linealidad), la imagen con orla se someterá sin más a la operación, siendo las dimensiones de la imagen de salida las mismas que las de entrada. Lo mismo se hará cuando la operación tenga lugar entre píxeles correspondientes de campos diferentes o de bandas de color diferentes.
- b) *En filtros espaciales de  $3 \times 3$ .* Supóngase que la imagen de entrada sin orla tiene la dimensión  $N_k$  (en una dimensión). La imagen de entrada con orla tendrá pues la dimensión  $N_k + 12$  y la imagen de salida con orla tendrá asimismo la dimensión  $N_k + 12$ . Se calcula en primer lugar el valor del bisel de salida (como en la anterior ecuación II-37) y se graba como mínimo en aquellos píxeles de bisel que no estén ocupados por la operación de imagen siguiente. A continuación, comenzando a un píxel de distancia del borde izquierdo

<sup>26</sup> El nivel 7 sólo es necesario para la construcción del mapa de crominancia.

de la imagen de entrada con orla, el núcleo de  $3 \times 3$  comienza a operar sobre la imagen de entrada sobregabando los valores del bisel de la imagen de salida, y deteniéndose a un píxel de distancia del borde derecho (o inferior) de la imagen (en el que se mantiene el valor de bisel original). El valor de bisel pregrabado hace innecesaria la operación del núcleo fuera de la imagen original (con orla) para calcular estos valores.

- c) *En el filtrado y reducción de la densidad de muestreo en REDUCE.* Dada una imagen de entrada con orla de dimensión  $N_k + 12$ , se le atribuye una matriz de salida de dimensión  $[N_k/2] + 12$ . El valor de bisel (calculado, por ejemplo, como en la ecuación II-37) se graba como mínimo en aquellos píxeles de bisel no ocupados en la operación de filtrado y reducción de densidad de muestreo siguientes. A continuación, se filtra la imagen de entrada con arreglo a b), pero se aplica el filtro a los píxeles  $-4, -2, 0, 2$  y  $4$  hasta agotar la imagen de entrada, grabándose los valores de salida en píxeles consecutivos  $-2, -1, 0, 1, 2, \dots$ , hasta que ya no quede sitio para ellos en la imagen de salida. Obsérvese que la posición del píxel 0 en la nueva imagen está a 7 píxeles del extremo izquierdo de la nueva imagen. La aplicación al último píxel del filtro lleva el píxel de entrada  $N_k + 3$  al píxel de salida  $[N_k/2] + 2$  cuando  $N_k$  es impar, y el píxel de entrada  $N_k + 4$  al píxel de salida  $[N_k/2] + 2$  cuando  $N_k$  es par. (En este contexto, se define el píxel de entrada como aquel que corresponde al centro del núcleo de 3 píxeles.)

A continuación se presentan cuatro ejemplos simplificados del procesamiento de marcos con REDUCE. En cada caso, los píxeles se etiquetan consecutivamente, entre corchetes y en negrita para la *imagen propiamente dicha*, y se subrayan para el bisel grabado con anterioridad.

EJEMPLO 1:  $N_k = 3$ . (Tamaño de entrada impar, tamaño de salida impar.)

Entrada:  $-6 \ -5 \ -4 \ -3 \ -2 \ -1 \ [0 \ 1 \ 2] \ 3 \ 4 \ 5 \ 6 \ 7 \ 8$   
 Salida:  $-6 \ -5 \ -4 \ -3$   $-2 \ -1 \ [0] \ 1 \ 2 \ 3 \ 4 \ 5 \ 6$

EJEMPLO 2:  $N_k = 4$ . (Tamaño de entrada par, tamaño de salida par.)

Entrada:  $-6 \ -5 \ -4 \ -3 \ -2 \ -1 \ [0 \ 1 \ 2 \ 3] \ 4 \ 5 \ 6 \ 7 \ 8 \ 9$   
 Salida:  $-6 \ -5 \ -4 \ -3$   $-2 \ -1 \ [0 \ 1] \ 2 \ 3 \ 4 \ 5 \ 6 \ 7$

EJEMPLO 3:  $N_k = 5$ . (Tamaño de entrada impar, tamaño de salida par.)

Entrada:  $-6 \ -5 \ -4 \ -3 \ -2 \ -1 \ [0 \ 1 \ 2 \ 3 \ 4] \ 5 \ 6 \ 7 \ 8 \ 9 \ 10$   
 Salida:  $-6 \ -5 \ -4 \ -3$   $-2 \ -1 \ [0 \ 1] \ 2 \ 3 \ 4 \ 5 \ 6 \ 7$

EJEMPLO 4:  $N_k = 6$ . (Tamaño de entrada par, tamaño de salida impar.)

Entrada:  $-6 \ -5 \ -4 \ -3 \ -2 \ -1 \ [0 \ 1 \ 2 \ 3 \ 4 \ 5] \ 6 \ 7 \ 8 \ 9 \ 10 \ 11$   
 Salida:  $-6 \ -5 \ -4 \ -3$   $-2 \ -1 \ [0 \ 1 \ 2] \ 3 \ 4 \ 5 \ 6 \ 7 \ 8$

- d) *Para el aumento de densidad de muestreo y filtrado de EXPAND.* Dada una imagen de entrada con orla del nivel  $k + 1$  con dimensión  $N_{k+1} + 12$ , y una matriz de salida del nivel  $k$  con dimensión  $N_k + 12$ , se le asigna una matriz de salida al nivel  $k$  con dimensión  $N_k + 12$ , y se inicializa 0. (Obsérvese que  $N_{k+1}$  se definió previamente como  $[N_k/2]$ .) A continuación, se insertan los píxeles de entrada  $-2, -1, 0, 1, \dots$  en los píxeles de salida  $-4, -2, 0, 2, 4, \dots$ . A continuación se efectúa la operación de filtrado de b) sobre la imagen resultante. Por último, se calcula el valor del bisel al nivel  $k$ , por ejemplo utilizando la ecuación II-37, y se graba en los tres píxeles exteriores a todos los lados de la imagen de salida. Obsérvese que la posición del píxel 0 en la nueva imagen está a siete píxeles del extremo izquierdo de ésta. La aplicación al último píxel del filtro lleva el píxel de entrada  $[N_k/2] + 2$  al píxel de salida  $N_k + 3$  si  $N_k$  es impar y el píxel de entrada  $[N_k/2] + 2$  al píxel de salida  $N_k + 4$  si  $N_k$  es par. (Una vez más, el píxel de entrada del filtro se define como aquel que corresponde al centro del núcleo de tres píxeles.)

A continuación se facilitan cuatro ejemplos simplificados del procesamiento de marcos con EXPAND. En cada caso, los píxeles se etiquetan consecutivamente, entre corchetes y en negrita para la *imagen propiamente dicha*, y se subrayan para el bisel grabado con posterioridad.

EJEMPLO 1:  $N_k = 3$ . (Tamaño de entrada impar, tamaño de salida impar.)

Entrada: -6 -5 -4 -3 -2 -1 [**0**] 1 2 3 4 5 6

Salida:                    -6 -5 -4 -3 -2 -1 [**0 1 2**] 3 4 5 6 7 8

EJEMPLO 2:  $N_k = 4$ . (Tamaño de entrada par, tamaño de salida par.)

Salida: -6 -5 -4 -3 -2 -1 [**0 1**] 2 3 4 5 6 7

Entrada:                    -6 -5 -4 -3 -2 -1 [**0 1 2 3**] 4 5 6 7 8 9

EJEMPLO 3:  $N_k = 5$ . (Tamaño de entrada par, tamaño de salida impar.)

Salida: -6 -5 -4 -3 -2 -1 [**0 1**] 2 3 4 5 6 7

Entrada:                    -6 -5 -4 -3 -2 -1 [**0 1 2 3 4**] 5 6 7 8 9 10

EJEMPLO 4:  $N_k = 6$ . (Tamaño de entrada impar, tamaño de salida par.)

Salida: -6 -5 -4 -3 -2 -1 [**0 1 2**] 3 4 5 6 7 8

Entrada:                    -6 -5 -4 -3 -2 -1 [**0 1 2 3 4 5**] 6 7 8 9 10 11

En estos ejemplos ilustrativos puede verse que la sobregrabación del bisel no produce ningún efecto cuando el proceso EXPAND se repite para varios niveles sucesivos.

## Apéndice II – Adjunto 1

### Bibliografía

- Recomendación UIT-R BT.500-11 (2002), *Metodología de evaluación subjetiva de la calidad de las imágenes de televisión*.
- USA Standards Committee T1\*, Technical Report T1.TR.73-2001: Video Normalization Methods Applicable to Objective Video Quality Metrics Utilizing a Full Reference Technique.

---

\* Las normas T1 son mantenidas por ATIS desde noviembre de 2003.

## Apéndice II – Adjunto 2

### Factores de prueba, tecnologías de codificación y aplicaciones

Para más información sobre los datos de estos cuadros véase el Informe final del VQEG (Material de referencia UIT-T). Todos los datos corresponden a los sistemas de 525 líneas.

**Cuadro II – Ad. 2.1 – Factores de prueba, tecnologías de codificación y aplicaciones para las que el método PQR alcanza la precisión especificada en II.1.3.4**

Velocidad binaria	Resolución	Método	Observaciones
2 Mbit/s	¾ resolución	mp@ml	Sólo hay reducción de la resolución horizontal
2 Mbit/s	¾ resolución	sp@ml	
4,5 Mbit/s		mp@ml	Con errores
3 Mbit/s		mp@ml	Con errores
4,5 Mbit/s		mp@ml	
3 Mbit/s		mp@ml	
4,5 Mbit/s		mp@ml	NTSC y/o PAL compuestos
6 Mbit/s		mp@ml	
8 Mbit/s		mp@ml	NTSC y/o PAL compuestos
8 & 4,5 Mbit/s		mp@ml	Dos códecs concatenados
19 Mbit/s – NTSC- 19 Mbit/s – NTSC- 12 Mbit/s		422p@ml	Tres generaciones de NTSC
50-50-... -50 Mbit/s		422p@ml	7ª generación con cuadro desplazado/I
19-19-12 Mbit/s		422p@ml	3ª generación
n/a		n/a	Betacam multigeneración con compensador de fallos de señal (4 ó 5, compuesto/componentes)

**Cuadro II – Ad. 2.2 – Factores de prueba, tecnologías de codificación y aplicaciones para las que el método PQR no alcanza la precisión especificada en II.1.3.4**

Velocidad binaria	Resolución	Método	Observaciones
1,5 Mbit/s	CIF	H.263	Pantalla completa
768 kbit/s	CIF	H.263	Pantalla completa
Otras			El método PQR especificado en el apéndice I no es adecuado para las aplicaciones de videoconferencia que repiten campos o no satisfacen los requisitos de latencia y retardo de las clases de vídeo. Además, el método PQR sólo es aplicable a los sistemas de transmisión de radiodifusión característicos con tasas de error muy bajas tales como los contemplados en las pruebas VQEG.

**Cuadro II – Ad. 2.3 – Secuencias de prueba utilizadas para determinar los factores de prueba, las tecnologías de codificación y aplicaciones para las que el método PQR alcanza la precisión especificada en II.1.3.4**

Secuencia	Características
Globo que revienta	Película, color saturado, movimiento
Nueva York 2	Efecto de enmascaramiento, movimiento
Móvil y calendario	Disponible en ambos formatos, color, movimiento
Animales que no lo son tanto	Color, síntesis, movimiento, corte de escena
El punto	Color, transparencia, movimiento en todas direcciones
Hojas de otoño	Color, paisaje, movimiento de zum, cascada
Fútbol	Color, movimiento
Velero	Prácticamente en reposo
Susi	Color de la piel
Tempestad	Color, movimiento

## Apéndice II – Adjunto 3

### Clasificación de errores

La clasificación de errores es una de las maneras de evaluar la eficacia de una métrica de calidad de vídeo (VQM). En este apéndice se explica el significado de la clasificación de errores, en términos de gráficas de la nota z subjetiva en función de la VQM delta descrita en el texto principal de esta Recomendación. Para la descripción siguiente, se utiliza la escala [0, 1] común tanto para la puntuación subjetiva como para la objetiva. En esta escala [0, 1] común, "0" representa la ausencia de degradación mientras que "1" representa la degradación máxima.

En cualquier prueba subjetiva se puede fijar un umbral  $\Delta z$ , que define cuándo dos puntos de datos (A, B) son estadísticamente equivalentes y cuándo son estadísticamente diferenciables<sup>27</sup>. Una vez realizado lo anterior, los resultados de la prueba subjetiva permiten clasificar cada par de puntos de datos (A, B) en una de las tres categorías siguientes:

$$\begin{aligned} \Delta z_{AB} < -\Delta z & \quad \rightarrow \text{A es mejor que B} & \quad \rightarrow \text{Bs} \\ -\Delta z \leq \Delta z_{AB} \leq \Delta z & \quad \rightarrow \text{A es igual que B} & \quad \rightarrow \text{Es} \\ \Delta z < \Delta z_{AB} & \quad \rightarrow \text{A es peor que B} & \quad \rightarrow \text{Ws} \end{aligned}$$

Las abreviaturas correspondientes a las tres categorías (Bs, Es y Ws) representan, respectivamente, la superioridad subjetiva, la equivalencia subjetiva y la inferioridad subjetiva.

<sup>27</sup> Los puntos de datos A y B representan en realidad conjuntos de observaciones de dos combinaciones SRC/HRC. Como se indica en el texto principal de esta Recomendación, la cantidad  $\Delta z_{AB}$  es la diferencia entre las medias de A y B ( $\hat{S}_{A\bullet} - \hat{S}_{B\bullet}$ ), dividida por la desviación típica inferida  $\sqrt{(V_A / N_A + V_B / N_B)}$ , siendo  $V_A$  la varianza de las puntuaciones de la situación A,  $N_A$  el número de observaciones de la situación A, etc.

Considérese ahora un umbral semejante para los valores VQM,  $\Delta_0$ :

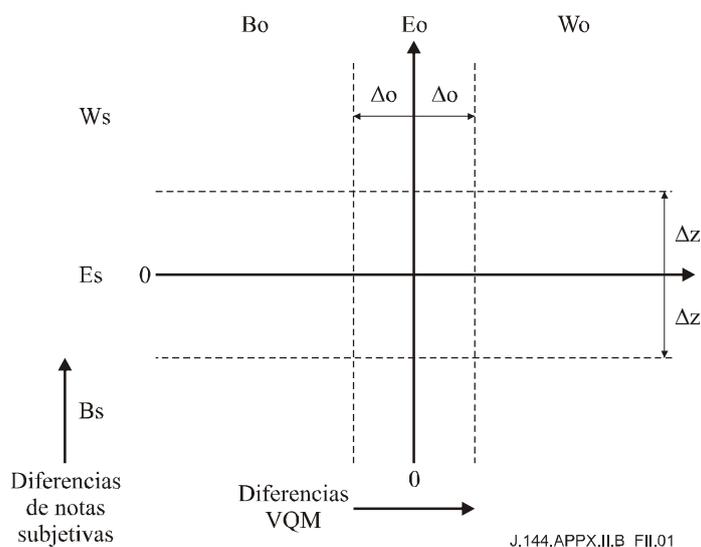
$$VQM(A) - VQM(B) < -\Delta_0 \quad \rightarrow \text{A es mejor que B} \quad \rightarrow B_0$$

$$-\Delta_0 \leq VQM(A) - VQM(B) \leq \Delta_0 \quad \rightarrow \text{A es igual que B} \quad \rightarrow E_0$$

$$\Delta_0 < VQM(A) - VQM(B) \quad \rightarrow \text{A es peor que B} \quad \rightarrow W_0$$

Las abreviaturas de estas tres categorías ( $B_0$ ,  $E_0$  y  $W_0$ ) representan la superioridad, equivalencia e inferioridad objetivas, respectivamente.

Dado que cada par de puntos de datos se somete a una clasificación triple en la prueba subjetiva y a otra clasificación triple independiente en la VQM, hay nueve resultados posibles. A continuación se representan gráficamente estos nueve espacios resultantes por las líneas de puntos en el espacio bidimensional de diferencias de notas subjetivas en función de la diferencia VQM:

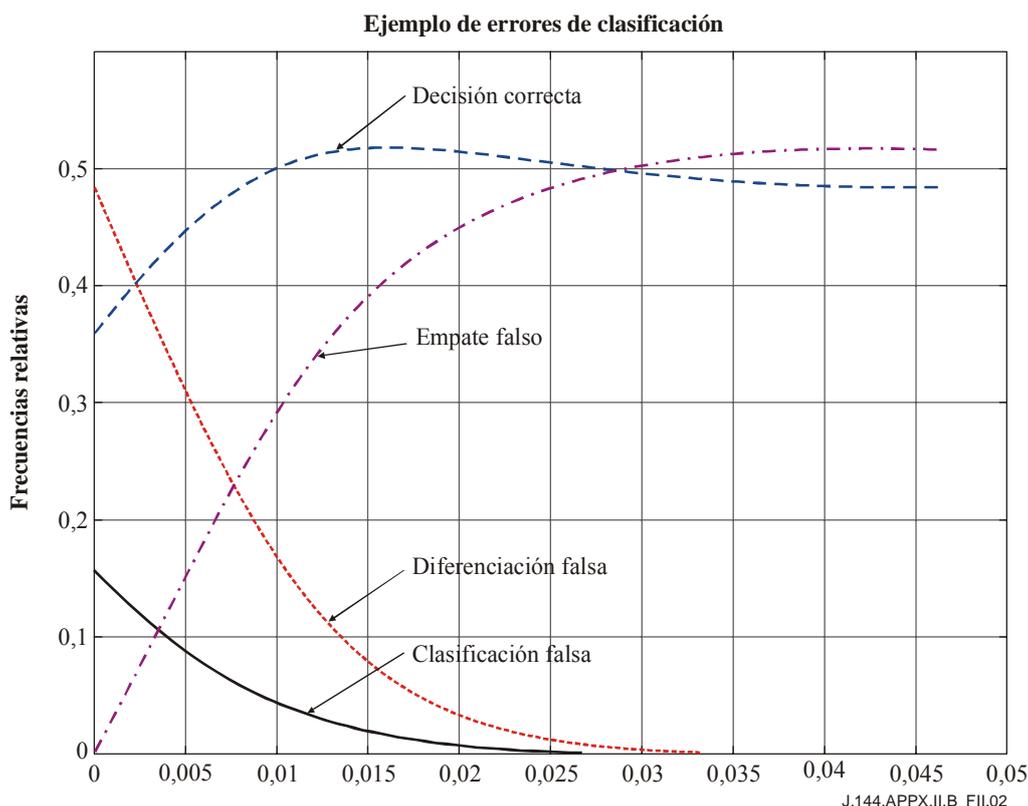


En el cuadro siguiente, se han etiquetado estos nueve resultados para poder responder a la pregunta "¿en qué se parecen la triple clasificación basada en VQM y la triple clasificación basada en la prueba subjetiva?".

	<b>Bs</b>	<b>Es</b>	<b>Ws</b>
<b>Wo</b>	Clasificación falsa	Diferenciación falsa	Decisión correcta
<b>Eo</b>	Empate falso	Decisión correcta	Empate falso
<b>Bo</b>	Decisión correcta	Diferenciación falsa	Clasificación falsa

Obsérvese que hay tres resultados en los que la clasificación VQM coincide con la de la prueba subjetiva. Estos tres resultados se han denominado "decisión correcta". Los seis resultados restantes corresponden a tres tipos diferentes de errores que pueden surgir cuando se utiliza una VQM. El empate falso es probablemente el error menos grave. Esto ocurre cuando la prueba subjetiva indica que dos puntos de datos son diferentes y la VQM indica que son idénticos. Una diferenciación falsa suele ser más grave. Esto ocurre cuando la prueba subjetiva indica que los dos puntos de datos coinciden y la VQM indica que son diferentes. La clasificación falsa sería normalmente el error más grave. En la clasificación falsa, la prueba subjetiva indica que A es mejor que B, y la VQM indica que B es mejor que A.

Para cualquier prueba subjetiva y cualquier VQM, se pueden formar todos los pares distintos posibles de puntos de datos y contar el número de pares de cada una de las cuatro categorías de resultados distintos: decisión correcta, empate falso, diferenciación falsa y clasificación falsa. A continuación se normaliza dividiendo por el número total de pares distintos y se anotan las frecuencias relativas de estas cuatro categorías de salida. Normalmente estos resultados serán función de  $\Delta s$  y de  $\Delta o$ . En el gráfico siguiente se muestra un ejemplo de resultados correspondientes a una VQM ficticia.  $\Delta z$  se ha seleccionado para obtener una confianza estimada del 95% en las clasificaciones subjetivas y  $\Delta o$  es el parámetro libre en el eje x de la gráfica.

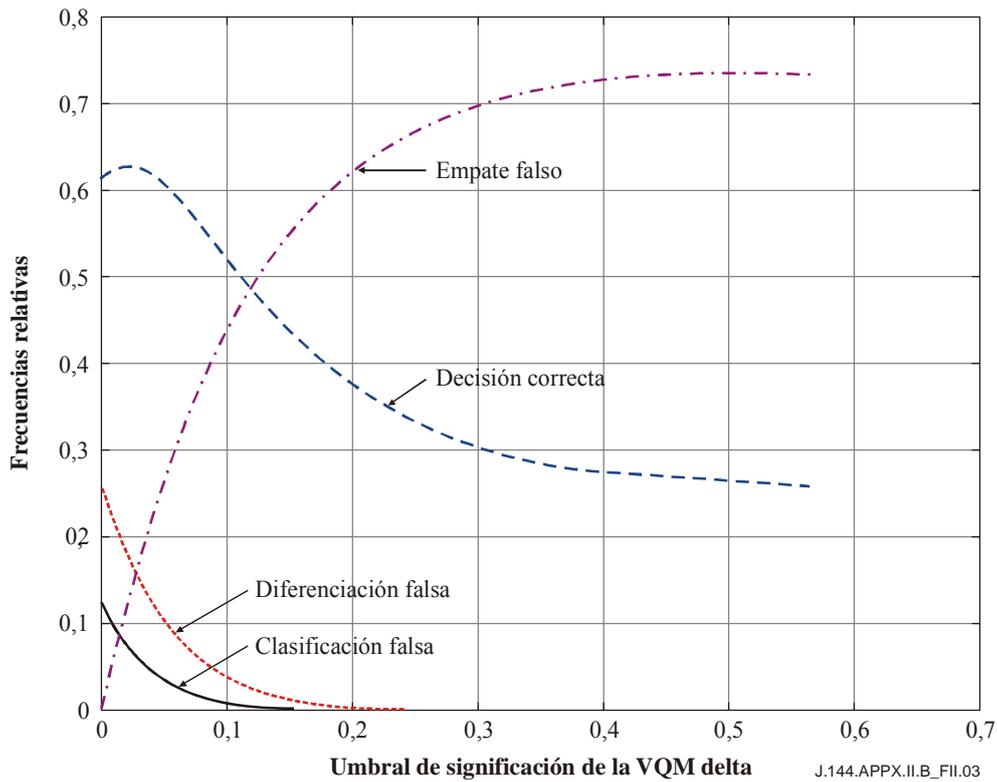


Obsérvese que conforme aumenta  $\Delta o$ , la VQM declara cada vez más pares de puntos de datos como equivalentes. Esto reduce el número de diferenciaciones falsas y de clasificaciones falsas, aunque aumenta el número de empates falsos. Conforme  $\Delta o$  se acerca a 0,05, la tasa de empates falsos tiende a 0,52. En este punto, la VQM declara la equivalencia de todos los pares, por lo que la VQM falla el 52% de las veces y acierta el 48%. Esto es coherente con el hecho de que se haya declarado la equivalencia de los pares de puntos de datos el 48% de las veces en la prueba subjetiva. Se podría utilizar una gráfica similar a ésta para seleccionar un valor adecuado de  $\Delta o$ . Por ejemplo, se podría escoger  $\Delta o$  de modo que se maximizara la probabilidad de adoptar decisiones correctas, o se podría seleccionar  $\Delta o$  para minimizar la suma ponderada de las frecuencias relativas de los errores.

En el código generador de la Figura anterior (parte del código MATLAB del anexo B), el umbral utilizado en la prueba subjetiva es `subj_th`. El umbral utilizado para  $\Delta VQM$  es `vqm_th` que queda como parámetro libre. El código permite dibujar la frecuencia de los tres tipos de errores diferentes y la ausencia de error, en función de `vqm_th`. Un valor óptimo de `vqm_th` podría ser aquel que maximizase la frecuencia de la ausencia de error, u otro que minimizase una suma de errores ponderados en costo. Obsérvese que, en general, es probable que los empates falsos sean los errores menos graves y que las diferenciaciones falsas sean los más graves, mientras que las clasificaciones falsas serán el peor tipo de error.

NOTA – Los nueve resultados y la cuadrícula de tres por tres en el espacio ( $\Delta VQM$ , puntuación  $Z$  subjetiva) constituyen el modo más lógico de describir este análisis. Esto supone la existencia de valores bipolares de  $\Delta VQM$ . No obstante, el código ya asume el valor absoluto de  $\Delta VQM$  (sustituyendo  $Z$  por  $-Z$  en todos los puntos cuyo valor de  $\Delta VQM$  sea negativo). Esto no modifica los cálculos matemáticos, sino que describe la situación de un modo más natural con seis resultados y una cuadrícula de 2 por 3. Dos de los resultados correctos (A mejor que B y A peor que B) se han superpuesto. Sigue habiendo dos resultados de empate falso, aunque solamente un resultado de diferenciación falsa y otro de clasificación falsa.

Para el método PQR especificado en el presente apéndice se aplica la siguiente clasificación de errores.





## SERIES DE RECOMENDACIONES DEL UIT-T

Serie A	Organización del trabajo del UIT-T
Serie D	Principios generales de tarificación
Serie E	Explotación general de la red, servicio telefónico, explotación del servicio y factores humanos
Serie F	Servicios de telecomunicación no telefónicos
Serie G	Sistemas y medios de transmisión, sistemas y redes digitales
Serie H	Sistemas audiovisuales y multimedios
Serie I	Red digital de servicios integrados
<b>Serie J</b>	<b>Redes de cable y transmisión de programas radiofónicos y televisivos, y de otras señales multimedios</b>
Serie K	Protección contra las interferencias
Serie L	Construcción, instalación y protección de los cables y otros elementos de planta exterior
Serie M	RGT y mantenimiento de redes: sistemas de transmisión, circuitos telefónicos, telegrafía, facsímil y circuitos arrendados internacionales
Serie N	Mantenimiento: circuitos internacionales para transmisiones radiofónicas y de televisión
Serie O	Especificaciones de los aparatos de medida
Serie P	Calidad de transmisión telefónica, instalaciones telefónicas y redes locales
Serie Q	Conmutación y señalización
Serie R	Transmisión telegráfica
Serie S	Equipos terminales para servicios de telegrafía
Serie T	Terminales para servicios de telemática
Serie U	Conmutación telegráfica
Serie V	Comunicación de datos por la red telefónica
Serie X	Redes de datos y comunicación entre sistemas abiertos
Serie Y	Infraestructura mundial de la información, aspectos del protocolo Internet y Redes de la próxima generación
Serie Z	Lenguajes y aspectos generales de soporte lógico para sistemas de telecomunicación