

RECOMENDACIÓN UIT-R BT.1788

Metodología para la evaluación subjetiva de la calidad de vídeo en aplicaciones multimedios

(Cuestión UIT-R 102/6)

(2006)

Cometido

Los sistemas de radiodifusión harán posible la distribución de aplicaciones de radiodifusión de datos y multimedios, incluyendo vídeo, audio, imagen fija, texto y gráficos. Esta Recomendación especifica métodos de evaluación subjetiva no interactivos para determinar la calidad de vídeo de aplicaciones multimedios.

La Asamblea de Radiocomunicaciones de la UIT,

considerando

- a) que se están introduciendo sistemas de radiodifusión digital en numerosos países;
- b) que los servicios de radiodifusión de datos y multimedios que comprenden vídeo, audio, imagen fija, texto, gráficos, etc. han sido introducidos, o se prevé introducirlos, utilizando sistemas de radiodifusión digital;
- c) que los servicios multimedios implicarán una infraestructura de radiodifusión caracterizada por el uso probable de receptores fijos y móviles velocidades de cuadro fijas y variables, diferentes formatos de imagen, códecs vídeo avanzados, pérdida de paquetes, etc.;
- d) que será necesario especificar los requisitos de calidad de funcionamiento y verificar si las soluciones técnicas consideradas son adecuadas para cada servicio teniendo presentes los requisitos de calidad de funcionamiento de dicho servicio;
- e) que la citada verificación implicará, principalmente, la evaluación subjetiva de la calidad de vídeo bajo condiciones controladas;
- f) que las metodologías de evaluación subjetiva especificadas en la Recomendación UIT-R BT.500 se pueden utilizar para aplicaciones multimedios;
- g) que también pueden usarse metodologías de evaluación subjetiva distintas de las especificadas en la Recomendación UIT-R BT.500;
- h) que la adopción de métodos normalizados es importante de cara al intercambio de información entre varios laboratorios,

recomienda

- 1 que los métodos generales de prueba, las escalas de apreciación y las condiciones de observación para la evaluación de la calidad de las imágenes descritas en el Anexo 1 se utilicen para las experiencias de laboratorio y, siempre que sea posible, para las evaluaciones prácticas de aplicaciones multimedios;
- 2 que se proporcione una descripción completa de las configuraciones y materiales de prueba, observadores y métodos en todos los informes de las pruebas;

3 que, para facilitar el intercambio de información entre los distintos laboratorios, los datos recopilados se procesen de acuerdo con las técnicas estadísticas indicadas en el Anexo 2.

NOTA 1 – La Comisión de Estudio 6 de Radiocomunicaciones debe proseguir con el desarrollo de una biblioteca de material de vídeo adecuado para la evaluación subjetiva de la calidad de vídeo en aplicaciones multimedios.

Anexo 1

Descripción de los métodos de evaluación

1 Introducción

Un gran número de países ha empezado a instalar sistemas de radiodifusión digital que permitirán la distribución de aplicaciones de radiodifusión de datos y multimedios que comprenden vídeo, audio, imagen fija, texto y gráficos.

Se necesitan métodos de evaluación subjetiva normalizados con objeto de especificar los requisitos de calidad de funcionamiento y de verificar si las soluciones técnicas consideradas para cada aplicación son las adecuadas. Las metodologías subjetivas son necesarias porque proporcionan mediciones que permiten a la industria anticiparse de manera más directa a las reacciones de los usuarios finales.

El sistema de radiodifusión necesario para distribuir aplicaciones multimedios es muy diferente del que se usa actualmente: se accede a la información a través de receptores fijos y/o móviles, la velocidad de cuadro puede ser fija o variable; hay una amplia gama de posibles tamaños de imagen (SQCIF a TVAD); el vídeo se asocia típicamente al audio, texto y/o sonido intercalados; el vídeo se puede procesar con códecs de vídeo avanzado; y la distancia de observación preferida depende en gran medida de la aplicación.

Los métodos de evaluación subjetiva especificados en la Recomendación UIT-R BT.500 deben aplicarse en este nuevo contexto. Asimismo, pueden llevarse a cabo investigaciones de sistemas multimedios mediante nuevas metodologías para ajustarse a los requisitos de usuario de las características del dominio multimedios.

En esta Recomendación se describen métodos de evaluación subjetiva no interactivos de la calidad de vídeo para aplicaciones multimedios. Dichos métodos pueden aplicarse para diferentes propósitos, que incluyen, sin ser exhaustivos: selección de algoritmos, clasificación de la calidad de funcionamiento de sistemas audiovisuales y evaluación del nivel de calidad de vídeo durante una conexión audiovisual.

Los términos y las definiciones asociados a esta Recomendación pueden consultarse en el Apéndice 3.

2 Características comunes

2.1 Condiciones de observación

En el Cuadro 1 se enumeran las condiciones de observación recomendadas. El tamaño y el tipo de visualización utilizados deben ser adecuados para la aplicación sobre la que se está investigando. Como deben emplearse varias tecnologías de visualización en las aplicaciones multimedia, es necesario comunicar toda información relevante sobre la visualización usada en la evaluación (por ejemplo, fabricante, modelo y especificaciones).

Cuando se haga uso de sistemas basados en PC para presentar las secuencias, deben señalarse asimismo las características de los sistemas (por ejemplo, la tarjeta de visualización de vídeo).

CUADRO 1
Condiciones de observación recomendadas empleadas
en evaluaciones de calidad de sistemas multimedia

Parámetro	Ajuste
Distancia de observación (Nota 1)	Limitado: 1-8 H Ilimitado: basado en las preferencias del espectador
Valor de la cresta de luminancia en la pantalla	70-250 cd/m ²
Relación entre la luminancia de la pantalla inactiva y la luminancia de cresta	≤ 0,05
Relación entre la luminancia de la pantalla, cuando se presenta únicamente el nivel de negro en una habitación totalmente oscura, y la correspondiente al blanco de cresta	≤ 0,1
Relación entre la luminancia de fondo detrás del monitor de imagen y la luminancia de cresta de la imagen (Nota 2)	≤ 0,2
Cromaticidad del fondo (Nota 3)	D ₆₅
Iluminación de fondo de la habitación (Nota 2)	≤ 20 lux

NOTA 1 – La distancia de observación depende en general de la aplicación.

NOTA 2 – Este valor indica un ajuste que permite la máxima detectabilidad de las distorsiones; para algunas aplicaciones se permiten valores superiores o vienen determinados por la aplicación.

NOTA 3 – Para monitores de PC, la cromaticidad del fondo debe aproximarse lo máximo posible a la de «punto blanco» de visualización.

En el Cuadro 2 se muestra un ejemplo del registro de datos para la configuración del sistema multimedia sometido a prueba.

Si las imágenes de prueba se obtienen por medio de una combinación decodificador-reproductor específica, hay que separar las imágenes de la interfaz gráfica patentada para lograr una visualización anónima. Ello es necesario para asegurar que la evaluación de la calidad no se ve influenciada por el conocimiento del entorno original.

Cuando los sistemas evaluados en una prueba utilizan un formato de imagen reducido, como CIF, SIF o QCIF, las secuencias deben mostrarse en una ventana de la pantalla de visualización. El color del fondo de la pantalla debe ser gris al 50%.

CUADRO 2

Configuración del sistema multimedia sometido a prueba

Parámetro	Especificación
Tipo de visualización	
Tamaño de visualización	
Tarjeta de visualización de vídeo	
Fabricante	
Modelo	
Información de la imagen	

2.2 Señales fuente

La señal fuente proporciona la imagen de referencia directamente, así como la entrada para el sistema que se está probando. La calidad de las secuencias fuente debe ser la mejor posible. Como referencia, la señal de vídeo debe grabarse en archivos multimedia usando yuv (formatos 4:2:2, 4:4:4) o RGB (24 ó 32 bits). Cuando el experimentador está interesado en comparar los resultados de diferentes laboratorios, es preciso utilizar un conjunto de secuencias común para eliminar fuentes adicionales de variación.

2.3 Selección de materiales de prueba

El número y tipo de escenas de prueba resultan críticos a la hora de interpretar los resultados de la evaluación subjetiva. Algunos procesos pueden provocar el mismo nivel de degradación para la mayor parte de las secuencias. En tales casos, los resultados obtenidos con un número pequeño de secuencias (por ejemplo, dos) deben proporcionar una evaluación representativa. No obstante, los nuevos sistemas a menudo tienen una repercusión que depende en gran medida de la escena o del contenido de la secuencia. En tales casos, el número y tipo de escenas de prueba debe escogerse de modo que pueda generalizarse de manera razonable la programación normal. Asimismo, el material debe seleccionarse de modo que sea «crítico, pero dentro de unos límites razonables» para el sistema que se está probando. «Dentro de unos límites razonables» implica que la escena podría todavía formar parte del contenido normal de la programación de televisión. Las características de percepción espacial y temporal de una escena pueden proporcionar una indicación útil de la complejidad de la misma. En el Apéndice 1 se presentan con más detalle las mediciones de características de percepción espacial y temporal.

2.4 Gama de condiciones y anclaje

Dado que la mayoría de los métodos de evaluación son sensibles a variaciones en la gama y distribución de las condiciones vistas, las sesiones de decisión deben incluir las gamas completas de los factores variados. Sin embargo, esto podría aproximarse mediante una gama más restringida presentando también algunas condiciones que corresponderían a los extremos de las escalas. Pueden representarse como ejemplos e identificarse como extremos (anclaje directo), o distribuirse a través de la sesión y no identificarse como extremos absolutos (anclaje indirecto). En la medida de lo posible, debe emplearse una gran gama de valores de calidad.

2.5 Observadores

El número de observadores tras el análisis debe ser de al menos 15; no expertos, en el sentido de que en su trabajo normal no están directamente interesados en la calidad de imagen y de que no son evaluadores con experiencia. Antes de una sesión, se debe analizar a los observadores para lograr (o corregir) la agudeza visual normal en el gráfico de Snellen o Landolt, y la visión en color normal, usando gráficos especialmente seleccionados (por ejemplo, Ishihara).

El número de evaluadores necesario depende de la sensibilidad y la fiabilidad del procedimiento de prueba adoptado, así como del tamaño anticipado del efecto buscado.

En los experimentos se debe incluir el mayor número posible de detalles acerca de las características de sus paneles de evaluación para facilitar una posterior investigación de este factor. Entre los datos que podrían proporcionarse figuran los de categoría laboral (por ejemplo, empleado de organización de radiodifusión, estudiante de universidad, empleado de oficina), sexo y edad.

2.6 Instrucciones para la evaluación

Debe familiarizarse detenidamente a los evaluadores con el método de evaluación, los tipos de degradaciones o los factores de calidad que probablemente se produzcan, la escala de apreciaciones, y la temporización. Las secuencias de entrenamiento que demuestran la gama y el tipo de degradaciones que van a evaluarse deben emplearse con escenas distintas a las utilizadas en las pruebas, pero de sensibilidad comparable.

2.7 Diseño experimental

Se deja a criterio del experimentador la selección de un diseño experimental con el que satisfacer los objetivos de coste y precisión específicos. Se recomienda incluir al menos dos reiteraciones (es decir, repeticiones de condiciones idénticas) en el experimento. Las reiteraciones permiten calcular la fiabilidad individual de cada sujeto y, si fuera necesario, descartar resultados no fiables de algunos sujetos. Además, las reiteraciones aseguran que los efectos del aprendizaje en una prueba quedan compensados en cierta medida. Se obtiene una mejora adicional en el tratamiento de los efectos del aprendizaje mediante unas pocas «presentaciones ficticias» al comienzo de cada sesión. Estas condiciones deben elegirse de modo que sean representativas de las presentaciones que se van a mostrar más tarde durante la sesión. Las presentaciones preliminares no se han de tener en cuenta en el análisis estadístico de los resultados de la prueba.

La duración de una sesión, es decir, de una serie de presentaciones, no debe ser más de media hora.

Cuando se prueban varias escenas o algoritmos, el orden de presentación de las escenas o algoritmos debe aleatorizarse. El orden aleatorio puede modificarse para garantizar que las mismas escenas o los mismos algoritmos no estén próximos temporalmente (es decir, consecutivamente).

3 Métodos de evaluación

Se puede examinar la calidad de funcionamiento del vídeo de los sistemas multimedios mediante las metodologías de la Recomendación UIT-R BT.500. En el § 3.1 de esta Recomendación se proporciona una lista de los métodos seleccionados.

En el § 3.2 se facilita una descripción de una metodología adicional denominada evaluación subjetiva de calidad de vídeo multimedios (SAMVIQ), que aprovecha las características del dominio multimedia y puede usarse para evaluar la calidad de funcionamiento de sistemas multimedios.

3.1 Metodologías de la Recomendación ITU-R BT.500

Para evaluar la calidad de vídeo en sistemas multimedios medios deben utilizarse las siguientes metodologías de la Recomendación UIT-R BT.500:

- Método de escala de degradación con doble estímulo (DSIS), como se describe en la Recomendación UIT-R BT.500, Anexo 1, § 4.
- Método de escala de calidad continua de doble estímulo (DSCQS), como se describe en la Recomendación UIT-R BT.500, Anexo 1, § 5.
- Métodos de estímulo único (SS), como se describe en la Recomendación UIT-T BT.500, Anexo 1, § 6.1.
- Métodos de comparación de estímulos (SC), como se describe en la Recomendación UIT-R BT.500, Anexo 1, § 6.2.
- Método de evaluación de calidad continua de estímulo único (SSCQE), como se describe en la Recomendación UIT-R BT.500, Anexo 1, § 6.3.

3.2 Evaluación subjetiva de la calidad de vídeo multimedios (SAMVIQ)

En este método, al espectador se le da acceso a varias versiones de una secuencia. Cuando el espectador ha calificado todas las versiones, puede accederse al siguiente contenido de la secuencia.

El espectador selecciona aleatoriamente las diferentes versiones por medio de una interfaz de ordenador gráfica, pudiendo detener, revisar y modificar como desee la nota otorgada a cada versión de una secuencia. Este método incluye una secuencia de referencia explícita (es decir, no procesada) y varias versiones de la misma secuencia que incluyen tanto secuencias procesadas como no procesadas (es decir, referencia oculta). Cada versión de una secuencia se visualiza por separado, clasificándose usando una escala de calidad continua similar a la empleada en el método DSCQS. Así, el método se asemeja mucho funcionalmente a un método de estímulo único con acceso aleatorio, pero un observador puede ver la referencia explícita siempre que quiera, haciendo que este método se asemeje a uno que utilice una referencia.

El método de evaluación de calidad SAMVIQ usa una escala de calidad continua que proporciona una medida de la calidad intrínseca de las secuencias de vídeo. Cada observador mueve un mecanismo deslizante en una escala graduada de 0 a 100, en la que están anotados 5 niveles de calidad dispuestos linealmente (excelente, bueno, normal, regular, malo).

La evaluación de calidad se lleva a cabo sucesivamente para cada escena (véase la Fig. 1, § 3.2.3), incluyendo una *referencia explícita*, una *referencia oculta* y *varios algoritmos*.

Para entender mejor el método, se definen a continuación las siguientes palabras específicas:

Escena: contenido audiovisual.

Secuencia: escena con procesamiento combinado o sin procesamiento.

Algoritmo: una o varias técnicas de procesamiento de imagen.

3.2.1 Referencia explícita, referencia oculta y algoritmos

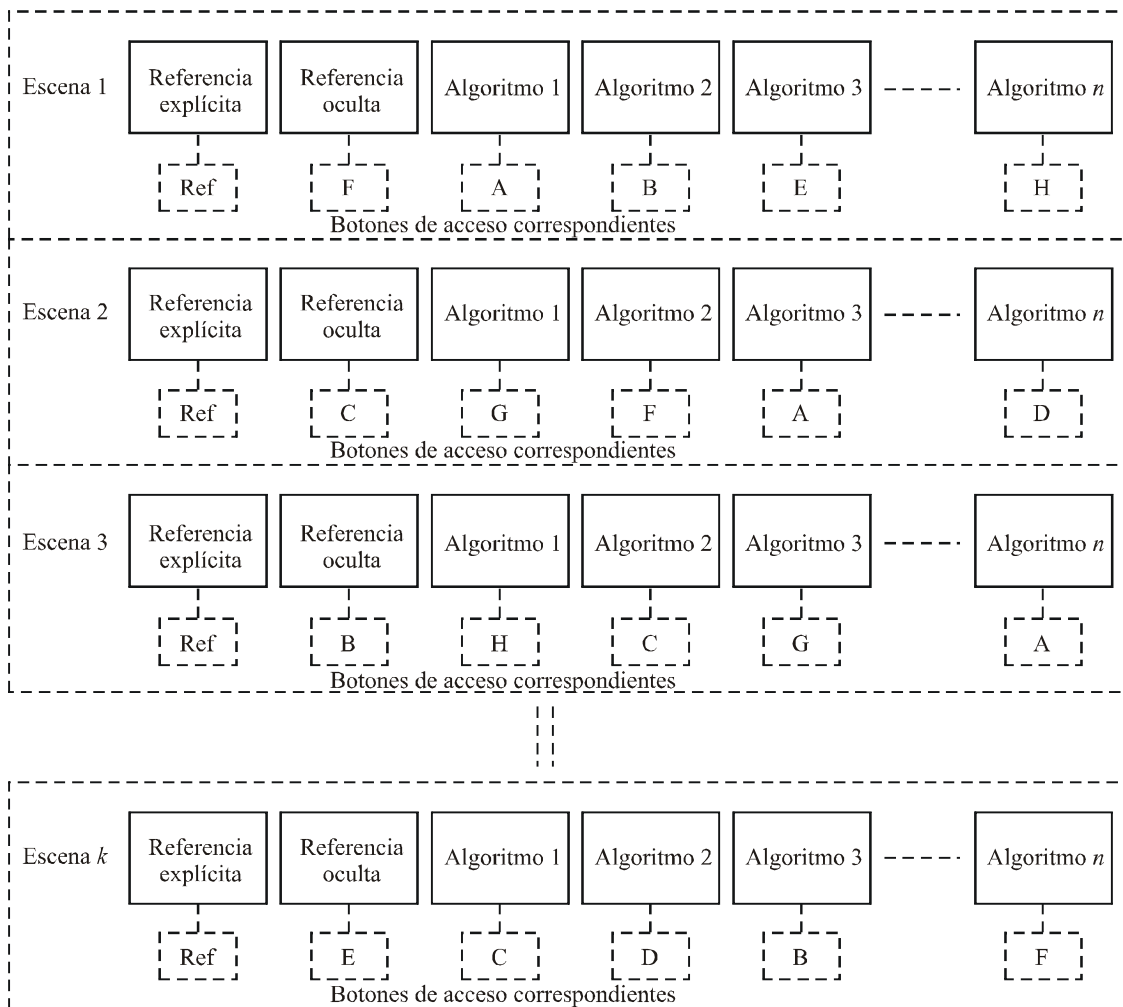
Un método de evaluación incluye normalmente anclajes de calidad para estabilizar los resultados. En el método SAMVIQ se consideran dos anclajes de alta calidad por las siguientes razones. Se han llevado a cabo varias pruebas que indican desviaciones típicas minimizadas de las notas usando una *referencia explícita* en lugar de una oculta, o de no usar referencia. Concretamente, para evaluar el rendimiento de los códecs, es mejor utilizar una referencia explícita con objeto de lograr la máxima fiabilidad de los resultados. Se añade también una *referencia oculta* para evaluar la calidad intrínseca de la referencia, en lugar de la referencia explícita, porque la presentación es anónima, al igual que las secuencias procesadas. El nombre explícito «referencia» tiene influencia sobre

aproximadamente el 30% de los observadores. Éstos dan la máxima nota posible (100) a la referencia explícita, siendo dicha nota totalmente diferente de la correspondiente a la referencia oculta. En particular, cuando no hay referencia disponible, la prueba sigue siendo posible, pero la desviación típica aumenta drásticamente.

El método SAMVIQ es adecuado para un contexto multimedios porque es posible combinar diferentes características de procesamiento de imagen, tales como tipo de códec, formato de imagen, velocidad binaria, actualización temporal, zooming, etc. La palabra *algoritmo* resume una de estas características, o combinación de las mismas.

FIGURA 1

Ejemplo de organización de prueba para el método SAMVIQ



3.2.2 Condiciones de la prueba

La variación de la criticidad de una escena es limitada porque el contenido homogéneo se escoge en función de las mismas reglas empleadas implícitamente por otras metodologías que proporcionan una nota global (por ejemplo, los métodos de estímulo único). Una máxima duración de visualización de secuencia de 10 ó 15 s es, por tanto, suficiente para lograr una nota de calidad estabilizada y fiable. Los decodificadores-reproductores patentados, o una copia de pantalla de su salida, deben utilizarse para mantener la calidad de funcionamiento de la visualización adecuada.

3.2.3 Organización de la prueba

- La prueba se lleva a cabo sucesivamente para cada escena, como se describe en la Fig. 1.
- Para la escena actual, es posible reproducir y otorgar una nota a cualquier secuencia en cualquier orden. Cada secuencia puede reproducirse y calificarse varias veces.
- De una escena a otra, el acceso secuencial se aleatoriza, evitando que los observadores intenten votar de un modo idéntico con arreglo a un orden establecido. De hecho, en una prueba el orden del algoritmo permanece invariable para simplificar el análisis y la presentación de los resultados. Sólo se aleatoriza el acceso correspondiente a partir de un idéntico botón.
- Para una primera visualización, la secuencia en curso debe reproducirse completamente antes de calificarse; de lo contrario, sería posible calificar y detenerse inmediatamente.
- Para probar la siguiente escena deben calificarse todas las secuencias de la escena en curso.
- Para concluir la prueba, deben calificarse todas las secuencias de cada escena.

El método SAMVIQ se implementa mediante programas informáticos. Además de los botones de acceso mostrados en la Fig. 1, se precisan los botones «reproducir», «detener», «escena siguiente» y «escena anterior», para permitir al espectador gestionar la presentación de las diferentes escenas (véase el Apéndice 2 como ejemplo). Cuando un espectador da una nota, ésta debe aparecer bajo el botón de acceso correspondiente a dicha escena. Cuando todas las versiones diferentes de una secuencia han sido calificadas, el espectador aún puede comparar las notas y modificar, si es necesario, sus valores. No es preciso revisar completamente la secuencia en curso porque ya han quedado patentes diferencias importantes durante la primera visualización.

Anexo 2

Presentación y análisis de datos

1 Información resumida

Se necesita información precisa sobre el entorno de la prueba para reproducir la misma, o para comparar los resultados de diferentes pruebas. Por tanto, se sugiere comunicar la información sobre el entorno de prueba como se describe en el Cuadro 3.

CUADRO 3

Información resumida de las pruebas

Nombre del método	
Tecnología de visualización	
Nombre de referencia de la visualización	
Valor de cresta de la luminancia (cd/m ²)	
Nivel de luminancia del negro (cd/m ²)	
Configuración del nivel de negro: PLUGE (umbral percibido de la distancia del nivel de negro a ultra negro = 8). De lo contrario, indica el valor umbral	
Nivel de luminancia de fondo (cd/m ²)	
Iluminación (lux)	
Distancia de observación: – No limitada: parte frontal de la visualización – Limitada: nH	
Tamaño de la visualización (diagonal en pulgadas)	
Relación de visualización entre anchura y altura	
Formato de visualización (N.º de columnas, N.º de líneas)	
Formato de entrada de imagen (N.º de columnas, N.º de líneas)	
Formato de salida de imagen ⁽¹⁾ (N.º de columnas, N.º de líneas)	
Temperatura de color blanco: D65 de lo contrario Coordenadas de color blanco (x, y)	
Número de observadores efectivos	

⁽¹⁾ Esta información se necesita cuando se procesa la imagen de entrada, por ejemplo al reescalarla, tras la visualización.

Las características de visualización pueden influir en los resultados de la prueba. Para monitores de pantalla plana se debe comunicar información adicional tal como la respuesta de luminancia (fidelidad gamma) y los colores primarios.

Las características de las secuencias de vídeo son importantes para diseñar una prueba o explicar sus resultados. Se sugiere comunicar las características espaciales y temporales como se describe en el Apéndice 1. Debe considerarse esta información en la colección de secuencias de prueba en la biblioteca de material vídeo adecuado para la evaluación subjetiva de la calidad de vídeo en aplicaciones multimedia.

2 Métodos de análisis

Los métodos de análisis son los descritos en la Recomendación UIT-R BT.500, Anexo 2, § 2.

3 Análisis de los observadores

Para los métodos enumerados en el Anexo 1, § 3.1, los procedimientos de análisis se describen en la Recomendación UIT-R BT.500, Anexo 2, § 2.3.

El análisis para SAMVIQ se describe en el siguiente punto. No obstante, este procedimiento podría usarse para los métodos SS, DSIS y DSCQS. El procedimiento es más sencillo de implementar que el correspondiente empleado en la Recomendación UIT-R BT.500 para dichos métodos.

3.1 Procedimiento de análisis SAMVIQ

Cada observador debe tener un método estable y coherente para votar una relativa degradación de calidad en cada escena y algoritmo. Los criterios de rechazo verifican el nivel de coherencia de las notas de un observador según la nota media de todos los observadores para una determinada sesión de pruebas. En el método SAMVIQ, como en el DSQCS, pueden considerarse todos los algoritmos (ocultos o referencia implícita, anclaje bajo, secuencias codificadas). El criterio de decisión se basa en una correlación de notas individuales con respecto a las notas medias correspondientes de todos los observadores de la prueba.

3.2 Correlación de Pearson

La relación entre la escala de calidad y la gama de notas de los observadores se supone lineal para aplicar la correlación de Pearson.

El principal objetivo es verificar mediante un método sencillo si las notas de un observador son coherentes con arreglo a las notas medias de todos los observadores en toda la prueba de la sesión. La referencia oculta se considera como un anclaje de alta calidad. Si se incluyen los anclajes bajos y altos, aumentan la nota de correlación, y a la inversa, los desplazamientos de correlación entre los observadores disminuyen.

$$r(x, y) = \frac{\sum_{i=1}^n x_i y_i - \left(\frac{\sum_{i=1}^n x_i}{n} \right) \left(\frac{\sum_{i=1}^n y_i}{n} \right)}{\sqrt{\left(\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i \right)^2}{n} \right) \left(\sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i \right)^2}{n} \right)}}$$

donde:

- x_i : nota media de todos los observadores para el trío (algoritmo, velocidad binaria, escena)
- y_i : nota individual de un observador para el mismo trío
- n : (número de algoritmo) \times (número de escenas)
- i : {número de códec, número de velocidad binaria, número de escena}.

3.3 Correlación de rango de Spearman

La correlación de rango de Spearman puede aplicarse incluso si la relación entre la escala de calidad y la gama de notas de los observadores no se supone lineal¹.

$$r(x, y) = \left[1 - \frac{6 \times \sum_{i=1}^n [R(x_i) - R(y_i)]^2}{n^3 - n} \right]$$

¹ Generalmente, los resultados de la correlación de Pearson están muy próximos a los de Spearman.

donde:

- x_i : nota media de todos los observadores para el trío (algoritmo, velocidad binaria, escena)
- y_i : nota individual de un observador para el mismo trío
- n : (número de algoritmo) \times (número de escenas)
- $R(x_i$ o $y_i)$: es el orden del rango
- i : {número de códec, número de velocidad binaria, número de escena}.

3.4 Criterios finales de rechazo para descartar a un observador de una prueba

Las correlaciones de rango de Spearman y de Pearson se llevan a cabo para descartar uno o varios observadores en función de las condiciones siguientes:

SI $[\text{media}(r) - \text{sdt}(r)] > \text{Umbral de correlación máxima (MCT)}$.

Umbral de rechazo = Umbral de correlación máxima (MCT).

EN OTRO CASO Umbral de rechazo = $[\text{media}(r) - \text{sdt}(r)]$.

SI $[r(\text{Observer}_i)] > \text{Umbral de rechazo}$.

ENTONCES observador «i» de la prueba no se descarta.

EN OTRO CASO observador «i» de la prueba se descarta.

donde:

$r = \min(\text{correlación de Pearson}, \text{correlación de rango de Spearman})$

$\text{mean}(r)$ promedio de las correlaciones de todos los observadores de una prueba

$\text{sdt}(r)$ desviación típica de las correlaciones de todos los observadores de una prueba

Umbral de correlación máxima (MCT) = 0,85.

El valor 0,85 de MCT es válido para los métodos SAMVIQ y DSCQS, en caso contrario debe considerarse el valor 0,7 para los métodos SS y DSIS.

Apéndice 1

Mediciones de información espacial y temporal

Las medidas de información espacial y temporal que figuran a continuación son de valor único para cada cuadro en una secuencia de prueba completa. Esto da lugar a una serie temporal de valores que por lo general tendrán un cierto grado de variación. Las medidas de información de percepción que figuran más adelante eliminan esta variabilidad con una función de máximo (máximo valor para la secuencia). La propia variabilidad se puede estudiar convenientemente, por ejemplo con muestras de información espacial-temporal cuadro a cuadro. La utilización de distribuciones de información a lo largo de una secuencia de prueba permite también una mejor evaluación de las escenas con cortes de escena

Información de percepción espacial (SI): Medida que generalmente indica el grado de detalle espacial de una imagen. Generalmente es mayor en escenas espacialmente más complejas. Esta información no constituye una medida de la entropía ni está asociada con la información definida en la teoría de la comunicación. La información de percepción espacial, SI, se basa en el filtro de Sobel. En primer lugar se filtra cada cuadro de vídeo (plano de luminancia) en un instante n (F_n) con el filtro de Sobel [$\text{Sobel}(F_n)$]. A continuación, se calcula la desviación típica de los píxeles ($std_{espacio}$) de cada cuadro filtrado con el filtro de Sobel. Esta operación se repite para cada cuadro de la secuencia de vídeo y da por resultado una serie temporal de información espacial de la escena. Se elige el máximo valor de la serie temporal (máx_{tiempo}) como representación del contenido de información espacial de la escena. Este proceso se puede representar en forma de ecuación como sigue:

$$SI = \text{máx}_{tiempo} \{std_{espacio} [\text{Sobel}(F_n)]\}$$

Información de percepción temporal (TI): Medida que generalmente indica la cantidad de cambios temporales de una secuencia de vídeo. Normalmente es mayor en secuencias de alta velocidad. Esta información no constituye una medida de la entropía ni está asociada con la información definida en la teoría de la comunicación.

La media de la información temporal, TI, se calcula como el máximo en el tiempo (máx_{tiempo}) de la desviación típica en el espacio ($std_{espacio}$) de $M_n(i, j)$ para todos i y j .

$$TI = \text{máx}_{tiempo} \{std_{espacio} [M_n(i, j)]\}$$

donde $M_n(i, j)$ es la diferencia entre píxeles en la misma posición en el cuadro, pero pertenecientes a dos cuadros consecutivos, es decir:

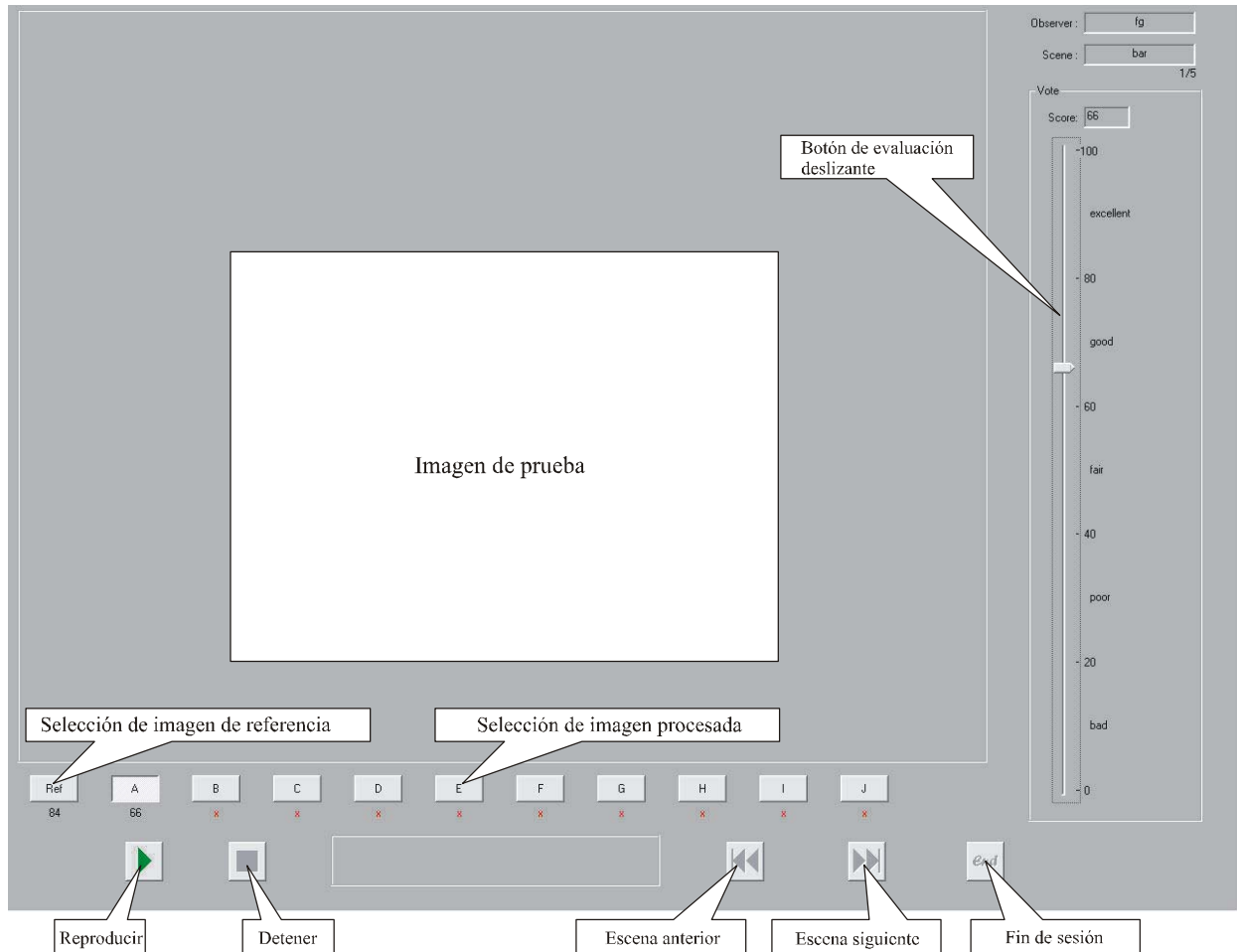
$$M_n(i, j) = F_n(i, j) - F_{n-1}(i, j)$$

donde $F_n(i, j)$ es el píxel de la i -ésima fila y j -ésima columna del n -ésimo cuadro en el tiempo.

NOTA 1 – Para escenas que contienen cortes, pueden proporcionarse dos valores: uno en el que el corte de escena se incluye en la medición de información temporal y otro en el que se excluye de la medición.

Apéndice 2

Ejemplo de interfaz para SAMVIQ



1788-02

Apéndice 3

Términos y definiciones

Algoritmo	Una o varias técnicas de procesamiento de imagen
AVI	Audio vídeo intercalado (<i>audio video interleaved</i>)
CCD	Dispositivo de acoplamiento de cargas (<i>charge coupled device</i>)
CI	Intervalo de confianza (<i>confidence interval</i>)
CIF	Formato intermedio común (<i>common intermediate format</i>) (formato de imagen definido en la Recomendación H.261 para vídeo-teléfono: 352 líneas × 288 píxeles)

CRT	Tubo de rayos catódicos (<i>cathode ray tube</i>)
DSCQS	Método de escala de calidad continua de doble estímulo (<i>double stimulus using a continuous quality scale method</i>)
DSIS	Método de escala de degradación con doble estímulo (<i>double stimulus using an impairment scale method</i>)
Escena	Contenido audiovisual
LCD	Pantalla de cristal líquido (<i>liquid crystal display</i>)
MOS	Nota media de opinión (<i>mean opinion score</i>)
PDP	Panel de visualización de plasma (<i>plasma display panel</i>)
PS	Segmento de programa (<i>programme segment</i>)
QCIF	Cuarto de CIF (<i>quarter CIF</i>) (formato de imagen definido en la Recomendación H.261 para videoteléfono: 176 líneas × 144 píxeles)
SAMVIQ	Evaluación subjetiva de calidad de vídeo multimedios (<i>subjective assessment of multimedia video quality</i>)
Secuencia	Escena con procesamiento combinado o sin procesamiento
SC	Método de comparación de estímulo (<i>stimulus comparison method</i>)
SI	Información de percepción espacial (<i>spatial perceptual information</i>)
SIF	Formato intermedio normalizado (<i>standard intermediate format</i>) [formatos de imagen definidos en ISO 11172 (MPEG-1): 352 líneas × 288 píxeles × 25 cuadros/s y 352 líneas × 240 píxeles × 30 cuadros/s]
S/N	Relación señal/ruido (<i>signal-to-noise ratio</i>)
SP	Presentación simultánea (<i>simultaneous presentation</i>)
SS	Método de estímulo único (<i>single stimulus method</i>)
SSCQE	Método de evaluación de calidad continua de estímulo único (<i>single stimulus using a continuous quality evaluation method</i>)
std	Desviación típica (<i>standard deviation</i>)
TI	Información de percepción temporal (<i>temporal perceptual information</i>)
TP	Presentación de prueba (<i>test presentation</i>)
TS	Sesión de prueba (<i>test session</i>)
VTR	Magnetoscopio (<i>video tape recorder</i>)
