Modelado de calidad percibida de video en Televisión Digital Abierta

Juan Pablo Garella Lena, *Member, IEEE,*

***Resumen*—In this master's thesis Video Quality Assessment state of the art is reached. Different full reference (FR), reduced reference (RR) and no reference (NR) objective video quality estimation models, previously proposed by the research group, by the academic community or standardized by international organizations, are evaluated and analyzed. Their performance is compared by contrasting their results with ones obtained through subjective video quality evaluations. The ISDB-Tb standard of free-to-air Digital Television (DTV) is deepened, from the point of view of the correct reception and visualization of the DTV signal. Finally, this work contains specific contributions in the area of objective video quality assessment with application in DTV. In this thesis it is proposed and evaluated a new method that combines FR and NR objective models to perform the perceived video quality prediction in real time for the DTV signal. This new approach is based on modern video quality evaluation techniques under a DTV monitoring system with receiver nodes geographically distributed in the coverage area of a DTV station.**

***Index Terms*—Digital Television (DTV), Quality of Experience (QoE), Video Quality Assessment (VQA).**

I. Introducción

En los servicios de comunicación audiovisual la calidad del contenido transmitido puede sufrir distintos tipos de degradaciones impactando en la calidad percibida de la imagen recibida. Particularmente en el caso de la Televisión Digital (TVD) estas degradaciones pueden asociarse a distintas etapas. En primera instancia se tiene el proceso de captura o adquisición de la señal de video y audio. Aquí, por ejemplo en el caso del video, la señal puede verse afectada con ruido o exposición, entre otros. La segunda etapa consiste del proceso de compresión digital de la señal realizado por el codificador, mientras la tercer etapa se da en la distribución y posterior transmisión de la señal. Aquí, por ejemplo, pueden ocurrir retardos o interferencia en el canal de radiofrecuencia. Por último, se encuentra la etapa de recepción y visualización por parte de los espectadores.

En general, al área de investigación que estudia la problemática de percepción de calidad de video se le denomina por sus siglas en inglés cómo VQA (*Video Quality Assessment*). Dependiendo del tipo de señal se pueden identificar otras ramas, cómo IQA (*Image Quality Assessment*) o evaluación de calidad de imágenes, extendiéndose a todo tipo de estimulo audiovisual o multimedia.

El procedimiento conocido más confiable para medir la calidad percibida de un contenido multimedia, por ejemplo una imagen o video, es la evaluación subjetiva. Ésta se debe realizar con un número determinado de personas, quienes opinan respecto de su percepción en ambientes controlados siguiendo recomendaciones de organismos internacionales. El resultado obtenido es la opinión media, denominada comúnmente MOS (*Mean Opinion Score*). También se utiliza el DMOS (*Degradation Mean Opinion*) asociado a la distorsión percibida entre el contenido multimedia degradado y el original. Estos procedimientos son costosos y complejos de realizar, en general implican varias sesiones de pruebas con un grupo considerable de personas, equipamiento, laboratorio y puesta a punto de las pruebas, no siendo un método practicable en aplicaciones de medición en tiempo real, ni replicables de manera permanente. Esto fundamenta la utilización de métodos automáticos y objetivos de estimación de calidad en los que no se precise la participación directa de usuarios que emitan su juicio respecto de la calidad percibida. Estos métodos objetivos generalmente se clasifican dentro de tres categorías: FR (*Full Reference*), RR (*Reduce Reference*), y NR (*No Reference*). Los métodos FR son aquellos que necesitan tanto de la señal original como de la degradada o procesada con el fin de computar la estimación de calidad. Los RR son aquellos que necesitan información parcial o reducida de la señal original y por último los métodos NR son aquellos que no necesitan de una referencia para estimar la calidad percibida de una señal degradada. Estos métodos (FR, RR y NR) se entrenan, verifican y validan contrastando sus resultados con aquellos obtenidos por medio de evaluaciones subjetivas y su desempeño generalmente se mide en términos de la correlación con el MOS, siendo éste el *"ground truth"*, es decir, el valor más confiable en cuanto a la medida de calidad percibida de un contenido multimedia. Cuanto mayor correlación tengan los resultados de un modelo objetivo con los resultados obtenidos por medio de evaluaciones subjetivas mejor desempeño tendrá dicho modelo. Para tal propósito existen bases de datos con clips de video. A menudo se utilizan clips de 10 segundos a varios minutos de duración con distintos tipos de distorsiones según el tipo de aplicación objetivo. De esta manera se pueden verificar y validar distintos modelos propuestos en la literatura.

En este trabajo se realiza un estudio del estado del arte en modelos objetivos de estimación de calidad de video así como de metodologías para la realización de evaluaciones subjetivas de la calidad. Tras la realización de evaluaciones subjetivas se presenta una novedosa base de datos generada con el fin de entrenar, verificar y validar modelos objetivos con aplicación en TVD. Por otro lado se diseña e implementa un novedoso método que combina modelos FR y NR con el fin de obtener una medida de la calidad percibida de video de la señal de TVD de manera objetiva y automática en el marco de un

This work was supported by Agencia Nacional de Investigación e Innovación (ANII), Montevideo, Uruguay.

Juan Pablo Garella Lena is with Instituto de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay, (e-mail: jpgarella@fing.edu.uy).

Manuscript received April 30, 2017; revised XXXX.

sistema de monitorización de la señal de TVD. Para ello se comparan los resultados obtenidos tras evaluar varios modelos con el fin de seleccionar el modelo que mejor se adapte a los requerimientos del sistema.

En lo que sigue se describe la motivación y el marco de trabajo de esta tesis de maestría.

I-A. Motivación

El despliegue de la Televisión Digital Abierta (TVD) en América Latina ha comenzado. Uruguay, así como la mayoría de los países en la región ha adoptado el estándar Japonés-Brasileño ISDB-Tb (Integrated Services for Digital Broadcasting, Terrestrial, Brazilian Version), también conocido como ISDB-T Internacional definido por la Asociación Brasileña de Normas Técnicas (ABNT) [1]. A lo largo de este documento se utilizará la sigla ISDB-T para hacer referencia a la versión brasileña y no a la original de la norma. El desarrollo de la TVD tiene implicancias técnicas y socio-políticas, promete una mejor calidad de imagen y una utilización más eficiente del espectro así como un aumento en la diversidad cultural del contenido. En este contexto la aceptación de dicha tecnología juega un papel clave y en gran medida depende de la calidad de experiencia (QoE, *Quality of Experience*) percibida por parte del público. Por su parte, para los reguladores y operadores también es importante conocer la calidad de servicio así como la calidad de experiencia que ofrecen a los usuarios. Poder contar con herramientas que permitan obtener esta información brinda a los operadores, en este caso los radiodifusores, la posibilidad de velar por la calidad de experiencia que ofrecen al público. A pesar de su importancia, aún hoy existen varias definiciones e interpretaciones de la QoE. En el contexto de las redes de telecomunicaciones en [1] se propone la siguiente definición:

"The degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state."

Por otro lado el sector de estandarización de ITU-T, *Focus Group on IPTV* propone la siguiente definición:

"The overall acceptability of an application or service, as perceived subjectively by the end user".

NOTES:

1- *Quality of Experience includes the complete end-to-end system effects (client, terminal, network, services, infrastructure, etc).*
2- *Overall acceptability may be influenced by user expectations and context.*

En vista de estas definiciones e interpretaciones medir la QoE no resulta una tarea sencilla, involucra múltiples factores que afectan la percepción global, además es una medida subjetiva y puede diferir de un usuario a otro, dependiendo del tipo de aplicación o servicio y de sus expectativas. Siguiendo esta línea de pensamiento se observa la necesidad de generar herramientas que permitan obtener una medida de la calidad percibida de video para la señal de TVD, siendo ésta uno de los principales factores que afectan la percepción global de la experiencia por parte del público.

I-B. Marco de trabajo

Acompañando la implantación de la TVD en el territorio uruguayo el gobierno ha incentivado a través de la Dirección Nacional de Telecomunicaciones y Servicios de Comunicación Audiovisual (DINATEL) [2] y de la Agencia Nacional de Investigación e Innovación (ANII) [3] la realización de proyectos vinculados a la temática. A continuación se describen algunos de estos. El primero, que dio el puntapié inicial para la realización de esta tesis, fue el proyecto titulado: "Indicadores de Calidad de Video" (VQI, *Video Quality Indicators*) [2], con apoyo del fondo ANII-FST-1-2012-1-8147, patrocinado por MIEM/DINATEL a través de ANII y ejecutado en conjunto por la Universidad de Montevideo y la Universidad de la República. El objetivo fue la generación de indicadores de calidad percibida de video, tanto subjetivos como objetivos, con aplicación en TVD. El segundo proyecto se tituló: "Sistema de Monitorización de la señal de TV Digital" (SMTVD), se realizó por medio del fondo ANII-FST-1-2013-1-13436, fue patrocinado por MIEM/DINATEL a través de ANII y ejecutado en conjunto por la Universidad del República y el Centro de Ensayos de Software (CES). En este proyecto se diseñó e implementó una plataforma tecnológica para la monitorización de la calidad de la señal de TV Digital ISDB-T. El sistema implementado permite medir un amplio rango de factores que afectan a la señal, desde potencia recibida, indicadores de artefactos (efecto de bloques, borrosidad, etc), hasta el resultado final, que es la calidad de experiencia por parte del usuario. Para ello se acotó el problema haciendo énfasis en la estimación de la calidad percibida de video y audio a través de modelos objetivos que intentan predecir el juicio de los usuarios. La finalidad del proyecto fue apoyar la regulación para mejorar la calidad de experiencia y si es posible contar con un conocimiento a priori en la predicción de la opinión de los televidentes. Este proyecto se subdividió en tres carriles, uno referente al desarrollo e implementación de la arquitectura central del sistema, otro referente al diseño, implementación y despliegue de los nodos de recepción y por último uno asociado al estudio de calidad de la señal.

Por otro lado cabe destacar el *Protocolo de Homologación de Receptores ISDB-Tb* [4] donde se describen las pruebas necesarias para homologar receptores de TVD *full-seg*, ya sea televisores (TV) o *Set Top Boxes* (STB), desde el punto de vista de la correcta recepción de video y audio. Actualmente este protocolo es utilizado por el Laboratorio Tecnológico

[1]Asociación Brasileña de Normas Técnicas, [en línea], http://www.abnt.org.br, último acceso: 7 de Febrero de 2016).

[2]Dirección Nacional de Telecomunicaciones y Servicios Comunicación Audiovisual, en línea, http://www.dinatel.gub.uy, último acceso: 7 de Febrero de 2016

[3]Agencia Nacional de Investigación e Innovación, [en línea],http://www.anii.org.uy, último acceso: 7 de Febrero de 2016

[4]Protocolo de Homologación de Receptores ISDB-Tb, [online], www.tvd.gub.uy/download.php?m=n&i=52, (Accessed: 20 April 2016).

del Uruguay (LATU) para ensayar receptores acorde a lo especificado en el Decreto 143/013 del Ministerio de Industria Energía y Minería, Poder Ejecutivo.

Es en estos proyectos que se enmarca el presente trabajo de tesis de maestría.

I-C. Organización del documento

En la sección II se comienza el trabajo explorando el estado del arte en materia de evaluación de calidad de video, en su modalidad objetiva y subjetiva, prestando especial atención en su aplicación en los sistemas de Televisión. En la sección III se describe la base de datos generada tras la realización de evaluaciones subjetivas de calidad de video. En la sección IV se describe la adaptación de un modelo objetivo paramétrico de estimación de calidad de video para la TVD y se contrastan sus resultados con los obtenidos en las pruebas subjetivas realizadas. En la sección V se presenta un novedoso sistema que permite efectuar una medición de calidad percibida de video en tiempo real para Televisión Digital Abierta en distintos puntos geográficos bajo un área de cobertura dada de una estación de Televisión. Se describe un método especialmente diseñado para dicha tarea que combina modelos objetivos con referencia completa (FR) y sin referencia (NR). En la sección VI se presentan las conclusiones del trabajo y líneas de investigación a futuro.

II. Evaluación de Calidad de Video

En esta sección se explora el estado del arte en evaluación de calidad de video, se analizan técnicas clásicas, modernas y estandarizaciones internacionales en lo que hace a la evaluación subjetiva y objetiva de la calidad. Se describen las metodologías comúnmente utilizadas para la realización de experimentos subjetivos. Por otro lado se analizan distintos modelos objetivos de estimación de calidad de video de tipo: FR, RR y NR. Se culmina la sección mencionando los distintos trabajos realizados por el Grupo de Expertos en Calidad de Video (*Video Quality Experts Group*, VQEG) [3] especialmente en materia de televisión.

II-A. Evaluación subjetiva de calidad de video

La calidad de video se puede medir de varias maneras. Tal y como se mencionó anteriormente la manera más confiable es mediante la realización de evaluaciones subjetivas. Estas evaluaciones son experimentos psicofísicos en los que un número determinado de sujetos de prueba califica un conjunto de estímulos (clips de videos). Estas pruebas son costosas en términos de tiempo (preparación y funcionamiento) y recursos humanos, por tanto deben ser diseñados con cautela. Por otro lado existen métodos objetivos para evaluar la calidad de video, por tanto las evaluaciones subjetivas son necesarias tanto para obtener las calificaciones de manera directa como para desarrollar y calibrar métodos objetivos.

Varios métodos de evaluación subjetiva de calidad de video son reconocidos y estandarizados en distintas recomendaciones por parte de ITU (*International Telecommunication Union*). Particularmente la recomendación ITU-R BT.500-13 [4] fue especialmente desarrollada para aplicaciones de Televisión en definición estándar (SD), mientras que la recomendación ITU-R BT.710-4 [5] extiende su uso para alta definición (HD). Por otro lado, la recomendación ITU-T P.910 [6] describe metodologías de evaluación subjetiva para aplicaciones multimedia. Recientemente, en 2014, fue aprobada la recomendación ITU-T P.913 [7]; ésta describe métodos para la evaluación subjetiva de la calidad de video, la calidad de audio, la calidad audiovisual de video por Internet y la calidad de distribución de televisión en cualquier entorno.

Cuadro I: Evaluación subjetiva de video: métodos reconocidos y estandarizados en distintas recomendaciones por parte de ITU:

Método	Recomendación
***DSIS** (Double Stimulus Impairment Scale)*	*ITU-R BT.500-13*
DSCQS (Double Stimulus Continuous Quality Scale)	*ITU-R BT.500-13*
SSCQE (Single Stimulus Continuous Quality Evaluation)	*ITU-R BT.500-13*
SDSCE (Simultaneous Double Stimulus for Continuous Evaluation)	ITU-R BT.500-13
ACR (Absolute Category Rating) también conocido cómo SS (Single Stimulus)	ITU-T P.910
DCR (Degradation Category Rating), Similar al método DSIS	ITU-T P.910
ACR-HR (Absolute Category Rating - Hidden Reference)	ITU-T P.910
PC (Pair Comparison)	ITU-T P.910

En la tabla I se mencionan los métodos más utilizados para evaluar calidad subjetiva de video estandarizados por ITU.

Se destaca el método de *Índices por categorías absolutas* (ACR, *Absolute Category Rating*) [6]. En este método, también denominado como *Single Stimulus* (SS), se presenta cada clip de video por separado. Los espectadores deben evaluar la calidad de cada clip de video en una escala discreta como la que se presenta en la tabla II. Una variante a este método se denomina por sus siglas en ingles como ACR-HR [4] (*Absolute Category Rating - Hidden Reference*) donde se incluye una versión sin degradaciones de cada una de las secuencias originales de prueba para actuar como referencia en la escala de evaluación de cada sujeto de prueba.

Cuadro II: Escala discreta de cinco niveles, (ACR)

5	Excelente
4	Buena
3	Aceptable
2	Mediocre
1	Mala

Por otro lado se tiene el método de escala de degradación con doble estímulo (DSIS, *Double Stimulus Impairment Scale*) [4]. En el método DSIS al espectador se le presenta primero el estímulo original (un clip con su calidad original), seguido por el mismo clip, pero degradado. En general estos clips tienen unos 10 segundos de duración. Después de cada par, al espectador se le pide votar el deterioro o degradación del segundo teniendo en cuenta el original. Para ello se utiliza la siguiente escala de cinco puntos, siendo: 5 - Imperceptible, 4 - Perceptible, 3- Ligeramente molesta, 2 - Molesta y 1 - Muy molesta. A este método también se lo conoce como Índices por Categorías de Degradación (DCR, *Degradation Category*

Rating) [4].

Con el correr de los años se han realizado numerosas evaluaciones subjetivas de calidad, ya sean imágenes, videos, audio o multimedia. Sin embargo se han presentado obstáculos para que dichas bases de datos sean utilizadas libremente por la comunidad científica con el fin de atacar problemas abiertos de investigación en la materia. Una de las razones por las cuales sucede esto es debido a los contratos legales relacionados a los productores, actores y propietarios que limitan cómo y dónde se puede utilizar el contenido. Con el objetivo de atacar este problema, según comentan sus creadores, se creó la biblioteca web llamada "The Consumer Digital Video Library"(CDVL) [8] [5]. La biblioteca contiene clips de video descomprimidos de alta calidad, públicos y gratis para su descarga, donde sus propietarios conceden el permiso de que sean utilizados con fines de educación, investigación y verificación de modelos objetivos en distintas aplicaciones. Por otro lado en [9] se pueden encontrar varias referencias a bases de datos públicas con videos calificados. Además existen esfuerzos para construir y utilizar bases de datos de videos en gran escala [10] [11].

II-B. Evaluación Objetiva de Calidad

Los modelos objetivos de estimación de calidad son procedimientos automáticos que utilizan algoritmos basados en características del contenido o parámetros de la red de trabajo con el fin de obtener una medida objetiva de calidad perceptual, ya sea para imágenes, videos, audio o incluso una combinación de estos, generando modelos audiovisuales o multimedia. Estos modelos son necesarios para asegurar la calidad en la entrega de contenido. Es el caso de la Televisión Digital Abierta, IPTV, o cualquier tipo de tecnología orientada a la emisión de contenido multimedia. Existen numerosos modelos objetivos propuestos a la fecha para evaluar la calidad de video [12]. Estos tratan de emular el comportamiento humano respecto de la percepción de calidad, en este sentido, su desempeño se mide en términos de correlación con los resultados obtenidos mediante evaluaciones subjetivas. Es decir, conforme estos modelos realicen una mejor estimación del MOS obtenido en pruebas subjetivas, mejor desempeño tendrán. Es por este motivo que los modelos objetivos son entrenados y calibrados a partir de evaluaciones subjetivas y validados con un conjunto independiente de videos etiquetados con sus respectivos resultados subjetivos. En [13], se realiza un estudio exhaustivo de su evolución, analizando sus características, ventajas e inconvenientes. Además se presentan aplicaciones de video basadas en QoE y se identifican posibles direcciones de investigación.

Los modelos objetivos se pueden jerarquizar de distinta manera, como ya se ha comentado, la más común o tradicional los subdivide en tres categorías: FR, RR y NR [14] . Esta clasificación se basa en el grado de utilización de una referencia para el cómputo de la estimación de calidad. Existen otros criterios de clasificación dependiendo del enfoque utilizado, aquellos que utilizan una aproximación psicofísica y los que toman una aproximación de ingeniería para atacar el problema [13]. La aproximación psicofísica esta principalmente fundamentada en la caracterización del sistema visual humano (HVS, *Human Visual System*), como el efecto de enmascaramiento, la función de sensibilidad de contraste (CSF, *Contrast Sensitivity Function*) y la adaptación al color y la iluminación, entre otros. Mientras que la aproximación de ingeniería se basa en la extracción y análisis de ciertos patrones de distorsión en el video, como artefactos de compresión (efecto de bloques, efectos de borde, etc). Por otro lado, en la literatura [15] se puede encontrar una tercer clasificación donde se encuentran los modelos paramétricos, que utilizan información del flujo de bits de información (*bitstream)*, de los cabezales o la carga útil (*payload*) de los paquetes para computar el MOS. Además se describen modelos híbridos que combinan distintas aproximaciones. Por último, adicionalmente a las clasificaciones anteriores, se denominan como modelos basados en pixeles o *pixel-based* a aquellos modelos con referencia completa que utilizan el contenido, valor de los pixeles, para computar el valor de similitud entre dos señales.

Una persona puede emitir su juicio subjetivo en cuanto a la calidad de un video en base a su experiencia, expectativas y factores que afecten su decisión sin necesidad de la señal original. Al igual que las personas, los modelos objetivos sin referencia NR no necesitan más que de la señal degradada para valorar objetivamente su calidad. Este tipo de modelos puede ser utilizado en aplicaciones de tiempo real de QoE, es el caso de la Televisión Digital Abierta en donde no es posible contar con la señal original. Se ha invertido mucho esfuerzo en generar este tipo de modelos. Por ejemplo, en [16] se propone un modelo de estimación de calidad con foco en imágenes con compresión JPEG. El modelo toma características del contenido vinculadas a degradaciones introducidas en el proceso de compresión, como los artefactos relacionados al efecto de bloques, con el fin de estimar la calidad percibida. Por otro lado se tienen los modelos que utilizan estadísticas de la red, por ejemplo, pérdida de paquetes, ancho de banda, parámetros de codificación: tasa de bits(*bitrate*), tasa de cuadros (*frame rate*), etc. En el caso de video-telefonía en [17] se propone un modelo paramétrico de estimación de calidad multimedia sobre redes IP en baja resolución. Sobre la base de este trabajo surge la recomendación ITU-T G.1070 en 2007 [18]. Por su parte en [19] José Joskowicz et al. presentan una revisión de modelos paramétricos publicados por distintos proponentes. Se describen y comparan sus desempeños utilizando un conjunto de clips de video, en diferentes escenarios de codificación y transmisión. Además, se presenta un modelo paramétrico que tiene en cuenta las degradaciones obtenidas en el proceso de codificación, así como también características extraídas del contenido del video y su transmisión sobre redes de paquetes IP.

En lo que sigue se describen brevemente los modelos objetivos de tipo FR y RR utilizados en este trabajo, se comienza por aquellos que hacen a la evaluación de calidad percibida en imágenes (IQA, *Images Quality Assessment*) y luego en video (VQA, *Video Quality Assessment*). Siendo los primeros extensibles a una secuencia de imágenes, permitiendo su aplicación para video.

[5]CDVL , [En línea], http//:www.cdvl.org

Relación Señal a Ruido de Pico (PSNR, Peak signal-to-noise ratio): Se comienza el estudio por uno de los modelos objetivos con referencia completa más utilizado en la rama del procesamiento de señales, el error cuadrático medio (MSE) y la relación señal a ruido de pico (PSNR). Se muestra su formulación en las ecuaciones 1 y 2 respectivamente, en donde, a modo de ejemplo, se utiliza una imagen de tamaño $M \times N$ en escala de gris.

$$\text{MSE}(f,g) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(f_{ij}-g_{ij})^2 \tag{1}$$

$$\text{PSNR}(f,g) = 10\log_{10}\left(\frac{L_{max}^2}{\text{MSE}(f,g)}\right) \tag{2}$$

Donde "f" es la imagen de referencia y "g" la degradada, L_{max} corresponde al máximo nivel de señal. Si la imagen fuera de 8 bits por pixel entonces L_{max} sería de $2^8-1=255$. Cabe destacar que su sencilla implementación se ve opacada por sus resultados, poco correlacionados con la percepción de calidad por parte del usuario [20]. Existen algunas variantes al tradicional PSNR, es el caso de los modelos PSNR-HVS [21] y PSNR-HVS-M [22] que toman en cuenta las características del Sistema Visual Humano (HVS, *Human Visual System*). Particularmente la función de sensibilidad de contraste (CSF, *Contrast Sensitivity Function*) y el efecto de enmascaramiento.

Índice de Similaridad Estructural (SSIM, Structural Similarity Index): En las últimas décadas, una gran cantidad de esfuerzo se invirtió en el desarrollo de nuevos métodos de evaluación de la calidad que se aprovechen de las características conocidas del HVS. De esta manera nace el SSIM propuesto en [23]. Este modelo se basa en la suposición de que el sistema visual humano está altamente adaptado para extraer información estructural del campo de visión. De ello se desprende que una medida del cambio de información estructural entre una imagen original y una degradada pueda proporcionar una buena aproximación a la percepción de distorsión. El índice se basa en la siguiente expresión [23]:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \tag{3}$$

En donde,

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{4}$$

$$\sigma_x = \left(\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)^2\right)^{\frac{1}{2}} \tag{5}$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y) \tag{6}$$

En evaluación de calidad de imágenes es útil aplicar el índice SSIM localmente en lugar de a nivel global. Para ello se aplica el índice de la ecuación 3 en una ventana móvil. En general se utiliza una ventana *Gaussiana* circular, simétrica y normalizada, de tamaño 11×11 [23]. Al resultado obtenido se le denomina mapa de distorsión. Luego con el objetivo de obtener el valor medio, denominado Mean SSIM o MSSIM, se realiza un promedio en el mapa de distorsión, según la siguiente expresión:

$$MSSIM(X,Y) = \frac{1}{M}\sum_{j=1}^{M} SSIM(x_j,y_j) \tag{7}$$

Donde X e Y representan dos imágenes, la original y la degradada, x_j e y_j es el contenido de la imagen en la ventana local con índice j, por último M es el número total de ventanas locales en la imagen.

En la figura 1 se presenta una imagen original y una degradada por compresión JPEG, el mapa de distorsión utilizando la suma de diferencias absolutas (SAD) entre la imagen degradada y la original y por último el mapa de distorsión utilizando el índice SSIM con una ventana *Gaussiana*. Se puede notar que la compresión JPEG utilizada para la compresión de la imagen original produce efectos molestos, artefactos, a lo largo de los límites del edificio y en el cielo. Además se puede observar que el mapa de distorsión asociado al índice SSIM captura con mayor detalle los artefactos de la imagen que el mapa de distorsión asociado al error absoluto. Por ejemplo, se puede observar que el efecto de bloques en el cielo está claramente resaltado. En ambos casos, un valor más brillante de pixel indica mejor calidad o fidelidad y un valor bajo, cercano a 0, fidelidad nula. Con el objetivo de generar una estimación de calidad percibida resta convertir el mapa de distorsión en un valor escalar, lo que en general se denomina *pool* o *pooling* de características. En el caso del Mean SSIM (MSSIM), tal y como alude su nombre, se realiza un promedio del mapa de distorsión. Sin embargo existen distintas variantes, algunas de las cuales se discuten en [24].

Con el paso de los años la métrica MSSIM, ha evolucionado. Por ejemplo, la métrica MS_SSIM [25] mejora el desempeño de su antecesora. En ésta se realiza un proceso de multi-escala. Este método es una manera conveniente de incorporar detalles de la imagen a diferentes resoluciones. Continuando con la evolución del índice SSIM existen distintas variantes. Por ejemplo, en [24] se proponen distintas alternativas para ponderar los valores del mapa de distorsión. Siguiendo esta linea se encuentra el modelo denominado IW-SSIM. Éste se describe en detalle en [26]. Por otra parte, sus autores destacan que al utilizar la misma metodología de extracción de información sobre el mapa de calidad generado por el modelo PSNR (MSE) se mejora el desempeño del PSNR tradicional [26]. De esta manera además generan una variante al PSNR denominada IW-PSNR.

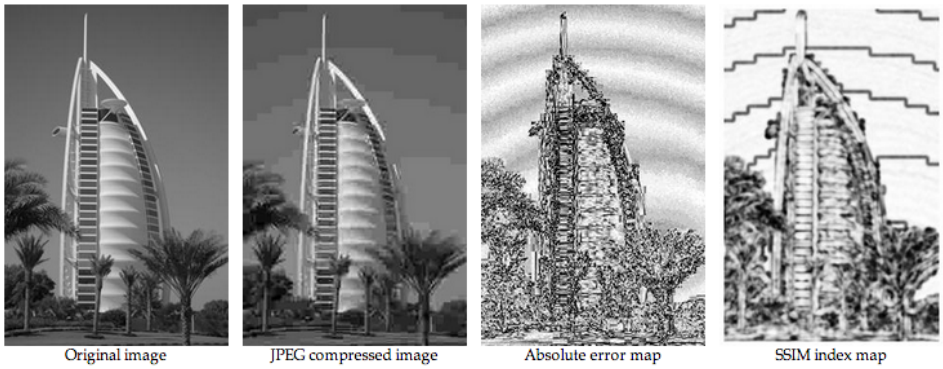


Figura 1: Imagen Original, Imagen degradada, Absolute Error Map, SSIM Index Map. Imagen extraída de [24].

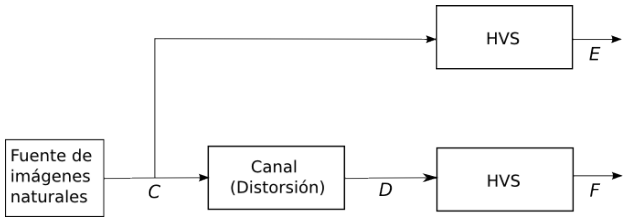


Figura 2: La información mutua entre la señal *C* y *E* cuantifica la información que el cerebro idealmente pude extraer de la imagen de referencia. Por otro lado la información mutua entre la señal *C* y *F* cuantifica lo que idealmente el cerebro puede extraer cognitivamente de la señal bajo ensayo.

Criterio de Fidelidad Visual de la Información (VIF, Visual Information Fidelity): El modelo VIF [27] es un modelo objetivo con referencia completa que ataca el problema de evaluación objetiva de la calidad como un problema de fidelidad visual de la información. En concreto, propone cuantificar la información presente en la imagen de referencia y cuanta de esta información puede extraerse de la imagen distorsionada. Combinando estas dos medidas de información se obtiene una medida visual de fidelidad de la información entre la imagen de referencia y su degradada. Con el fin de lograr este objetivo combina conceptos relacionados al campo de teoría de la información, como la información mutua, con el modelado de escenas naturales (NSS, *Natural Scene Statistics*) y el sistema visual humano (HVS). Las imágenes y videos capturados del entorno provienen de una clase común, denominada generalmente como escenas naturales. Éstas forman un pequeño subconjunto en el espacio de todas las posibles señales. Con el paso de los años se han desarrollado sofisticados modelos que capturan estadísticas de estas escenas. La gran mayoría de las degradaciones producidas en el mundo real distorsionan estas estadísticas y hacen que se le quite naturalidad a las imágenes o videos. En [28] se realiza una revisión de modelos basados en NSS. Específicamente en este método se modelan las imágenes naturales en el dominio de las wavelets utilizando el modelo GSM (*Gaussian Scale Mixtures*) presentado en [29].

En la figura 2 se describe a modo ilustrativo el esquema de trabajo del modelo VIF. La señal de referencia *C* se modela como una fuente estocástica "natural"que posteriormente pasa por el sistema visual humano para ser procesada por el cerebro. Se cuantifica la información de la imagen de referencia como la información mutua entre la señal de entrada *C* y de salida *E* del bloque asociado al HVS. Esta es la información que el cerebro idealmente puede extraer de manera cognitiva de la imagen original. Luego se cuantifica la misma medida pero en presencia de un canal con distorsión, es decir, se toma la información mutua entre la señal de referencia *C* y la señal *F* a la salida del bloque HVS previamente distorsionada por el canal. La relación o *ratio* entre estas dos medidas es el resultado del modelo VIF, según se presenta en la expresión 8.

$$VIF = \frac{\sum_{j=0}^{S} \sum_{i=1}^{M_j} I(c_{i,j}; f_{i,j})}{\sum_{j=0}^{S} \sum_{i=1}^{M_j} I(c_{i,j}; e_{i,j})} \tag{8}$$

Donde, S representa el número de sub-bandas en la descomposición por *wavelets*, M_j representa el número de bloques en la sub-banda j, $I(x, y)$ representa la información mutua entre x e y, c representa un vector (bloque) en una región especifica de la imagen de referencia, e representa la percepción del bloque c por un observador humano (con ruido blanco aditivo n), finalmente f representa la percepción del bloque c con distorsión.

Este modelo tiene propiedades interesantes, al igual que el modelo SSIM el mínimo valor que puede tomar es 0, este caso sucede cuando el cerebro no puede extraer información de la señal distorsionada, es decir la información mutua entre las señales C y F es nula. Por otro lado es igual a la unidad cuando extrae el mismo nivel de información para la imagen de referencia y su degradada. Por último la propiedad que lo hace distinto al resto de los modelos objetivos expuestos es que si se produce, por ejemplo, una mejora de contraste en la imagen distorsionada que no agregue ruido, el VIF resultará en un valor mayor a la unidad, indicando que la imagen degradada tiene mayor calidad que la de referencia. Esto tiene sentido dado que en general un mejora lineal del contraste tiende a mejorar la calidad percibida en una imagen.

Los autores de este modelo además proveen una implementación alternativa y simple del punto de vista de costo computacional. Ésta, parte de un proceso multi-escala en el dominio del pixel y le denominan VIFp.

Básicamente toma la ecuación 8, donde S se convierte en el número de escalas tomadas de la imagen en el dominio del pixel (bajo un proceso de filtrado gaussiano y submuestreo), mientras M_j denota el número de bloques en la escala j. Acorde a sus autores, el mayor costo computacional que tiene el modelo basado en *wavelets* proviene, justamente del cómputo en la descomposición por *wavelets* y el cálculo de los parámetros asociados al modelo de distorsión del canal. Por otro lado, como desventaja, a costa de disminuir el costo computacional la implementación VIFp varía levemente el desempeño del modelo original VIF.

VQM General NTIA : El modelo VQM General [30] fue propuesto por NTIA (National Telecommunications and Information Administration) y estandarizado por ANSI (American National Standards Institute) en 2003 e incluido en dos recomendaciones: ITU-T J.144 [31] e ITU-R BT.1683 [32]. Este modelo, acorde con lo estipulado en [31] utiliza parámetros objetivos con el fin de medir efectos perceptuales de una amplia gama de degradaciones, tales como borrosidad, efecto de bloques, movimiento entrecortado o innatural, ruido (tanto en señal de luminancia como en la de crominancia) y bloques con errores (asociados a errores en la transmisión digital). Con el objetivo de realizar la estimación de calidad el modelo toma una combinación lineal de determinados parámetros de video. De esta manera produce valores de salida que van de cero (sin degradación percibida) a uno (máxima degradación percibida). Ocasionalmente, puede tomar valores mayores que uno en las escenas de video extremadamente distorsionadas.

VQM LowBw NTIA : Una de las variantes propuestas al modelo VQM *General* es el modelo VQM *Lowbw* [33] de referencia reducida, también desarrollado por NTIA. Éste utiliza técnicas de extracción de características similares a las del modelo VQM *General* pero fue entrenado y ajustado con clips de video en diversos tamaños de pantallas, diversos *frame rates* y con degradaciones de pérdida de paquetes. Su desempeño fue evaluado en el proyecto RR/NR TV del VQEG [34]. Lleva el nombre "Lowbw" ya que requiere de muy bajo ancho de banda (10 Kbits/s) para enviar la información de referencia del video original para su comparación con el video degradado. Este modelo fue estandarizado en 2010 en la Recomendación ITU-T J.249 [35].

II-C. Video Quality Experts Group

El VQEG [3] (Video Quality Experts Group) lleva a cabo un gran trabajo sistemático y objetivo de comparación de modelos. Su objetivo es proporcionar evidencia a organismos internacionales de estandarización acerca del desempeño o rendimiento de diversos modelos propuestos, a los efectos de definir modelos estándar y objetivos de calidad percibida de video. En la tabla III se detallan los proyectos finalizados por el VQEG, su respectiva fecha de fin, la aplicación y el tipo de modelo objetivo (FR/RR/NR) evaluado. Además se indica, en caso de corresponder, las recomendaciones y estandarización realizadas por ITU en base a los resultados obtenidos para cada proyecto.

Por ejemplo, el proyecto *FRTV Phase I* fue el primer experimento de validación realizado por VQEG, donde se examinaron modelos objetivos FR, NR con el objetivo de predecir la calidad de la televisión de definición estándar (625 líneas y 525 líneas). Los modelos se presentaron en 1999 y el informe final de VQEG fue aprobado en junio del 2000 [36]. Como resultado, todos los modelos sin referencia (NR) fueron descartados por su bajo desempeño. Por otro lado, ITU decidió que el desempeño de los modelos FR no era suficiente para justificar su estandarización.

Sobre la base de esta experiencia y con el conocimiento adquirido se realizó el proyecto *FRTV Phase II*, donde nuevamente se analizaron modelos objetivos FR y NR con el objetivo de predecir la calidad de la televisión de definición estándar (625 líneas y 525 líneas). Los modelos se presentaron en 2002 y el informe final de VQEG fue aprobado en Agosto del 2003. Nuevamente todos los modelos NR fueron descartados, pero en esta instancia, ITU decide que el desempeño de algunos de los modelos FR propuestos, es el caso del modelo VQM *General* antes mencionado, era suficiente para justificar una estandarización.

Continuando con el trabajo del VQEG, en el proyecto *Multimedia Phase I* se utilizaron varias resoluciones (VGA, CIF, QCIF), además con el objetivo de comparar el desempeño de los modelos propuestos se utilizaron secuencias de video con degradaciones originadas en el proceso de compresión (artefactos) y errores en transmisión. Nuevamente se descartaron los modelos NR propuestos dado su bajo desempeño. Como resultado del proyecto, ITU realizó varias estandarizaciones, incluyéndose el modelo PSNR como implementación de referencia de mínimo desempeño (ver tabla III).

En el proyecto *RRNR-TV* se examinó modelos RR y NR con el objetivo de predecir la calidad de la televisión de definición estándar (625 líneas y 525 líneas). Todos los modelos NR fueron descartados. Como resultado, ITU consideró que el desempeño de algunos modelos RR justificaba su estandarización, es el caso del modelo VQM *LowBw*, por lo tanto se generó la recomendación ITU-T Rec. J.249, (2010) [35]. En el proyecto *HDTV Phase I* se examinaron modelos objetivos FR, RR y NR con el objetivo de predecir la calidad de la televisión de alta definición. Todos los modelos NR fueron descartados. Como resultado, ITU consideró que el desempeño de algunos modelos FR y RR justificaba su estandarización (ver tabla III). Por último en el proyecto *Hybrid Perceptual/Bitstream* se validó modelos objetivos de estimación de calidad que utilizan ambos, el video procesado o degradado y el bitstream o flujo de bits de información. Se utilizó resoluciones de video en WVGA, VGA y HD. Se examinaron modelos NR y modelos híbridos FR, RR y NR. El informe final fue aprobado en Julio de 2014, a partir de los resultados obtenidos se generó la recomendación ITU-T Rec. J.343 (2014) [37].

Actualmente el VQEG tiene en curso los siguientes proyectos[6]:

- *Audiovisual HD* (AVHD)
- *HDR/WGG* (High Dynamic Range Video / Wide Color

[6]Proyectos en curso, VQEG, [En linea], http://www.its.bldrdoc.gov/vqeg/projects-home.aspx

Cuadro III: Proyectos finalizados por el VQEG

Proyecto	Fin	Definición	ITU - Estandarización / Recomendaciones		
			FR	RR	NR
FRTV Phase I	Junio, 2000	SD	-	No aplica	-
FRTV Phase II	Agosto, 2003	SD	**(VQM General)** ITU-T Rec. J.144 ITU-R Rec. BT.1683	No aplica	-
Multimedia Phase I	Setiembre, 2008	Varias (VGA, CIF, QCIF)	ITU-T Rec. J.247 ITU-T Rec.J.246 ITU-R Rec. BT.1866 ITU-R Rec. BT.1867 **(PSNR)** ITU-T Rec. J.340	No aplica	-
RRNR-TV	Junio, 2009	SD	No aplica	**(VQM LowBw)** ITU-T Rec. J.249	-
HDTV Phase I	Junio,2010	HD	ITU-T Rec. J.341	ITU-T J.342	-
Hybrid Perceptual / Bitstream	Setiembre, 2014	Varias (WVGA, VGA, HD)	ITU-T Rec. J.343.5 ITU-T Rec. J.343.6	ITU-T Rec. J.343.3 ITU-T Rec. J.343.4	ITU-T Rec. J.343.1 ITU-T Rec. J.343.2

Gamut)
- *IMG* (Immersive Media Group)
- *JEG-Hybrid*
- *MOAVI* (Monitoring of Audio Visual Quality by Key Indicators)
- *PsyPhyQA* (Psycho-Physiological Quality Assessment)
- *QART* (Quality Assessment for Recognition and Task-based multimedia applications)
- *Ultra HD*
- *VIME* (Video and Image Models for consumer content Evaluation)
- *VLQA* (Visually Lossless Quality Analysis)

Por ejemplo, en el marco del proyecto VIME, el VQEG a liberado una herramienta de tipo *open source* denominada VIQET (VQEG Image Quality Evaluation Tool), disponible en linea [7]. Es una herramienta de evaluación de calidad de imágenes a partir de modelos objetivos sin referencia y ya se encuentran disponibles los primeros ejecutables. Por su parte, el proyecto MOAVI se centra en estudiar y modelar indicadores claves ("key indicators") de calidad audiovisual [38], [39], entre los cuales se incluyen el efecto de bloques, la borrosidad, el congelamiento de imágenes, efecto de entrelazado, entre otros.

En cuanto a los sistemas de televisión en agosto de 2012, ITU ha estandarizado la Recomendación ITU-R BT.2026, [40] que establece directrices para la implantación de sistemas de medición y supervisión objetivas para la cadena de distribución de programas de SDTV y HDTV. Estas directrices indican la necesidad de medir la calidad de video, y enviar las medidas obtenidas a puntos de supervisión. A los efectos de medir la calidad, se recomienda el uso de modelos objetivos del tipo RR, en particular los estandarizados en las recomendaciones ITU-R BT.1885 [41] y ITU-R BT.1908 [42], para SDTV y HDTV, respectivamente. Estos modelos objetivos fueron resultado directo de las actividades del VQEG. Dado que son modelos RR, requieren tener cierta información acerca del video original para poder estimar la calidad. Sin embargo, no está estandarizada la manera de enviar esta información del video original, y en los hechos, no es enviada por los *broadcasters*. Esto lleva a que hoy en día en la práctica, los modelos apropiados para realizar la estimación en línea de la calidad audiovisual sean los modelos NR, por ejemplo los paramétricos.

[7] www.GitHub.com/VIQET

III. Evaluación subjetiva de la calidad percibida de video en TV Digital

En esta sección se presenta la base de datos generada en el marco del proyecto VQI [2] tras la recolección de 8900 calificaciones individuales en evaluaciones subjetivas de calidad percibida de video con aplicación en Televisión Digital Abierta. Esta base se describe en [43] y encuentra a disposición [2] para la comunidad de investigación en QoE, con el fin de entrenar, verifica y validar modelos objetivos de evaluación de calidad de video. El sistema informático utilizado para la realización de las pruebas subjetivas se presenta en [44] y se encuentra disponible para la comunidad de QoE en [2]. Básicamente, es un sistema basado en software que automatiza las pruebas subjetivas para medición de la calidad de video. Con este sistema varios espectadores pueden participar de la misma sesión de evaluación, reduciendo así el tiempo total consumido por los ensayos. Además no necesita dispositivos o equipamiento especiales.

La base de datos se compone de dos conjuntos de clips de video sometidos a pruebas subjetivas de calidad, de aquí en más *Set I* y *Set II*. Cada conjunto cuenta con un centenar de videos en alta definición (HD) y otro centenar en definición estándar (SD), totalizando 400 clips de video etiquetados con su correspondiente evaluación subjetiva (MOS). Se pueden identificar en total 4 subconjuntos, de ahora en más *Set I HD*, *Set I SD*, *Set II HD* y *Set II SD*.

El *Set I* contiene video clips de laboratorio con degradaciones controladas y fue concebido para entrenar y verificar modelos objetivos. Mientras que el *Set II* cumple la función de conjunto de validación independiente, al utilizar video clips grabados de aire como videos de prueba.

Se utilizó la recomendación ITU-BT.500-13 [4] que describe la metodología a seguir para realizar pruebas subjetivas y además incluye posibles maneras de desplegar el contenido a los evaluadores. Siguiendo esta recomendación se utilizó la metodología "Single Stimulus" (SS), incluyendo las secuencias de referencia. En particular el método conocido como "Absolute Category Rating with Hidden Reference"

(ACR-HR) descrito en [6]. Cada video clip se presentó de manera independiente al evaluador, se utilizaron clips de 9 a 12 segundos de duración. Los clips originales se presentaron sin observaciones especiales. Además en cada una de las sesiones realizadas acorde a lo estipulado en la recomendación, el orden de presentación de los videos fue de manera aleatoria.

III-1. Set I: Con el fin de ser utilizado para entrenar y verificar modelos de estimación de calidad de video, el *Set I* cubre un amplio rango de situaciones en las que los modelos serán aplicados. Como criterio de selección de los videos se consideró el tipo de contenido, la actividad espacial y la actividad temporal. Un total de cinco clips de video con excelente calidad en resolución HD (1920×1080 pixeles) fueron utilizados como videos fuente. Éstos, fueron obtenidos de la biblioteca *CDVL* [8] y de la base de datos *IRCCyN IVC1080i Video Quality Database* [45]. Cada una de estas secuencias de video fue codificada utilizando compresión H.264/AVC y una sintaxis de tipo MPEG-2 TS (*Transport Stream*). Para ello se utilizó el popular programa *FFmpeg* [46]. Los parámetros de codificación utilizados se describen en la tabla IV.

Para el *Set I HD* se utilizaron cinco tasas de bits distintas, cubriendo así un rango de uso real en la transmisión de Televisión Digital. Esto constituye un conjunto de 25 clips de video en alta resolución degradados por artefactos de codificación. Para el *Set I SD* se utilizaron cuatro tasas de bits distintas, obteniendo 20 clips de video degradados con artefactos de codificación.

Con el ánimo de incluir errores en transmisión, un grupo reducido de los videos obtenidos del *Set I HD* y *Set I SD* fue sometido a un procedimiento de extracción individual de paquetes TS, simulando los errores en transmisiones de TVD. El tipo de degradaciones introducido es basado en patrones de pérdida reconocidos en grabaciones de transmisiones de TVD con baja recepción en el receptor [47]. Los distintos patrones de pérdidas de paquetes TS aplicados se pueden observar en la tabla V.

Cuadro IV: Parámetros de codificación H.264

Parámetros	**SD** (720×576)	**HD** (1920×1080)
Perfil	Main	High
Nivel	3.1	4.1
Largo del GoP (Group of Pictures)	33	33
Cuadros B consecutivos	2	2
Modo	CBR	CBR
Bitrates (Mbps)	0.7, 1.5, 2.8, 4.0	3.5, 5, 7.5, 10, 14
Tipo de barrido	Progresivo	Progresivo
Frame rate	50 fps	50 fps
Sintaxis de Flujo	MPEG2-TS	MPEG2-TS

En la figura 3 se muestran dos vistas previas de cuadros afectados por pérdida de paquetes para cada una de las cinco secuencias de origen seleccionadas.

III-2. Set II: El *Set II* fue concebido como un conjunto de clips de video independiente para realizar validación de modelos, al igual que el *Set I* se compone de un centenar

Cuadro V: Patrones de pérdida de paquetes TS probados

Patrones de pérdida	**Porcentaje de pérdida de paquetes TS**
Sin pérdida de paquetes	0 % a lo largo del clip de video
Una Ráfaga	0.1 % dentro de la ráfaga; 0 % fuera de la ráfaga
Una Ráfaga	10 % dentro del la ráfaga; 0 % fuera del ráfaga
Dos ráfagas	0.1 % dentro de la ráfaga; 0 % fuera de la ráfaga
Dos ráfagas	10 % dentro de la ráfaga; 0 % fuera de la ráfaga
Pérdida Uniforme	0.3 % a lo largo del clip de video

Figura 3: Vista previa de cuadros afectados por simulación de errores en transmisión en cada uno de los contenidos seleccionados

de video clips en HD y otro centenar en SD. Para ello se realizaron grabaciones de aire de unos 10 segundos de duración sobre dos canales de TVD ubicados en Montevideo, Uruguay. Estos se encontraban emitiendo señales de test en fase de pruebas. Además cabe resaltar que utilizaban distintos codificadores H.264. Cientos de grabaciones fueron registradas. El conjunto final se seleccionó tomando en cuenta la actividad temporal, espacial, cantidad de cortes de escena, contenido, artefactos de codificación, y patrones de pérdida de paquetes. En cuanto a los parámetros de codificación se notó estructuras de GOP variables incluso dentro de un mismo canal. A modo de ejemplo las señales grabadas contenían las siguientes secuencias de cuadros (GOP):

- IPBBBPBBBPBBI (largo GOP = 12, hasta 3 B seguidos)
- IBBPBBPBBPBBI (largo GOP = 12, hasta 2 B seguidos)

- IPPPPPPPPPPI (largo GOP = 12, hasta 0 B seguidos)

Además se notó que la cantidad de *slices* por cuadro era variable según el canal. Se registraron casos de 1 slice y 6 slices por cuadro.

Tras analizar los resultados obtenidos en la base de datos se puedo observar que en los clips de video con pérdidas uniformes la calidad percibida se ve fuertemente afectada, mientras que, si no hay pérdida de paquetes TS la calidad en general tiende a disminuir con la caída del bitrate. Por otro lado las ráfagas de errores tienen resultados de calidad variables. En algunos casos se puede ver que la calidad no se ve afectada y en otros disminuye de manera significativa. En las siguientes secciones se explotan las características de estos clips de video y los resultados de las evaluaciones subjetivas (MOS) para conseguir estimar su calidad percibida por medio de un modelo paramétrico objetivo.

IV. Aplicación de un modelo paramétrico de estimación de calidad de video en TV Digital

En esta sección se describe la adaptación y aplicación de un modelo paramétrico NR de estimación de calidad de video, generado en el marco del proyecto VQI. Se tomó como base el modelo presentado en [19]. Cabe destacar que este modelo fue pensado para funcionar sobre redes IP. Dado que las características de transmisión de la TVD son diferentes a las redes de paquetes IP, en primera instancia se debió adaptar al sistema ISDB-T de TVD. Para ello se tuvo en cuenta la influencia de las degradaciones asociadas a las pérdidas de paquetes en transmisión de TVD y también el uso de la resolución HD, no prevista en el modelo inicial.

En la ecuación 9 se presenta la formulación general del modelo,

$$MOSp = 1 + IcIp \tag{9}$$

Donde $MOSp$ es la predicción del MOS, Ic es la calidad dada por el material en sí mismo y el proceso de codificación e Ip representa la degradación introducida por el proceso de transmisión (pérdida de paquetes TS). $MOSp$ toma valores entre 1 y 5; conforme aumenta mayor será la estimación de la calidad percibida, Ic varía entre 0 y 4 e Ip entre 0 y 1. El máximo valor de $MOSp$ se obtiene cuando Ic es 4 e Ip es 1 y el más bajo cuando uno o los dos son 0. Si el material por sí mismo tiene buena calidad, y el proceso de codificación es bueno, Ic será cercano a 4. En cambio, en función de la degradación de la calidad asociada al proceso de codificación (por ejemplo bajo bitrate) este valor tenderá a 0. Sucede algo similar con Ip. Si las pérdidas de paquetes TS producen una distorsión excesiva de la calidad del video, Ip será cercano a 0. Mientras que si no hay pérdida de paquetes, será 1. En este último caso si no hay pérdida de paquetes la calidad percibida únicamente dependerá de Ic según la siguiente expresión:

$$MOSp = 1 + Ic \tag{10}$$

A continuación se presentan los modelos paramétricos asociados a Ic e Ip.

IV-A. Ic - Degradación asociada al proceso de codificación

En el modelo paramétrico seleccionado, Ic se calcula según la siguiente ecuación [19]:

$$Ic_{Pred} = v_3 \left(1 - \frac{1}{1 + \left(\frac{ab}{v_4}\right)^{v_5}} \right) \tag{11}$$

Donde,

$$v_3 = 4 + 4(f_{max} - f)(k_1 s + k_2 e^{-k_3(f_{max}-f)ab}) \tag{12}$$

$$v_4 = c_1 s^{c_2} + c_3 \tag{13}$$

$$v_5 = c_4 s^{c_5} + c_6 \tag{14}$$

Donde Ic_{Pred} es la predicción de Ic, la variable b representa al bitrate, a es un parámetro cuyo valor depende de la resolución de la imagen, f es el frame rate, siendo f_{max} igual al máximo *frame rate* utilizado y $c_1, c_2, c_3, c_4, c_5, c_6, k_1, k_2, k_3$ coeficientes del modelo. Por último s es una característica asociada a la complejidad de codificación del contenido del clip de video, denominada SAD (*Sum of Absolute Differences*). Esta característica se puede ver como el valor medio tras realizar el promedio de los valores de pixel asociados a cada uno de los bloques residuales obtenidos para cada uno de los vectores de movimientos computados en el correr de una secuencia de cuadros. Esta característica acorde a los resultados obtenidos en [48] muestra una alta correlación con los parámetros v_4 y v_5, de allí es que surgen las ecuaciones 13, 14.

La idea intuitiva detrás de la ecuación 11 es que ésta modele la componente de calidad Ic en función del bitrate, la resolución y el contenido en sí mismo del clip de video. Por ejemplo, para una resolución fija y un contenido dado, es de esperar que si se incrementa el bitrate aumente la componente de calidad Ic. En el límite ésta satura en un determinado valor (en este caso, en el valor asociado a v_3). Por su parte si uno mantiene el bitrate y el clip de video fijos y baja la resolución del clip es de esperar que la componente de calidad aumente. Esto basado en que se disminuye la cantidad de información a ser codificada, por lo tanto si se mantiene la tasa de información se obtendrá una mayor calidad. Por otro lado si se mantiene el bitrate y la resolución pero se decide variar el clip de video se puede tener en consecuencia un aumento o disminución en la complejidad de codificación. Esto depende del tipo de contenido utilizado. Este cambio es captado por la variable s (SAD) y puede resultar en una disminución o aumento de la componente de calidad Ic, respectivamente.

Con el fin de calibrar este modelo se utilizó como base de entrenamiento solo aquellos clips de video del conjunto *Set I* sin afectación de calidad por errores en transmisión, es decir solo aquellos con degradaciones asociadas al proceso de codificación. En total se utilizaron 25 clips en HD y 20 clips en SD. En la figura 4 se muestra la dispersión entre Ic_{Pred} e Ic sobre los clips de video de entrenamiento seleccionados. Como resultado se obtuvo una correlación de Pearson (PLCC) de 0.8385 y un RMSE de 0.3356.

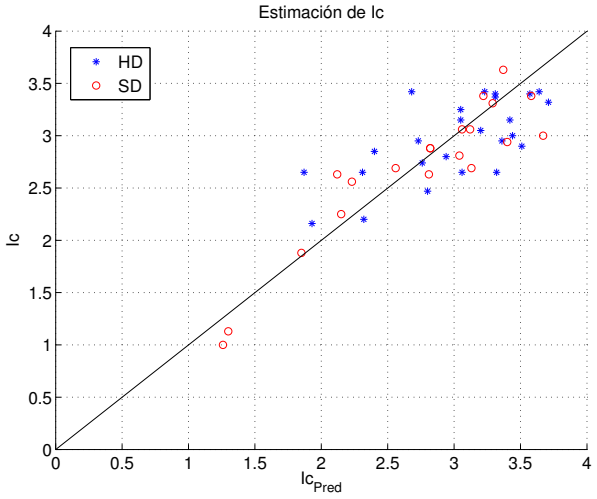


Figura 4: Estimación de Ic sobre la base de datos de entrenamiento

IV-B. Ip - Degradación de calidad asociada a errores en transmisión

Cada uno de los paquetes que se pierden debido a errores en el canal pueden contener información relacionada a cuadros, en el caso de H.264/AVC de *slices* tipo I, P o B. En H.264/AVC los cuadros conformados por uno o varios *slices* de tipo I se utilizan para decodificar cuadros con *slices* de tipo P o B. Por lo tanto se puede decir que no todos los paquetes tienen la misma influencia en la degradación de calidad percibida. Por ejemplo, en caso de que se pierdan macrobloques de un slice de tipo I la propagación del error en el *GOP* será mayor que al perderse información de un slice tipo B. Además los *slices* de tipo P son utilizadas para decodificar *slices* de tipo B en el GOP. Finalmente una degradación en *slices* de tipo B no afecta al resto de los cuadros. Tomando como base esta propagación de errores en el GOP, con el fin de obtener una estimación de Ip en [47] se propone un modelo que utiliza las ecuaciones 15 y 16. En esta tesis se denomina a este modelo de estimación de Ip por el nombre $Ip_{Inicial}$.

$$Ip_{Inicial} = \frac{1}{1 + kP_w} \tag{15}$$

$$P_w = x_1 I + x_2 P + B \tag{16}$$

Donde I es el porcentaje de *slices* de tipo I directamente afectados por la pérdida de paquetes respecto al total de *slices* de tipo I en el clip de video, P es el porcentaje de *slices* de tipo P afectados respecto al total de *slices* de tipo P, y B es el porcentaje de *slices* de tipo B afectados respecto al total de cuadros B. Por último x_1, x_2 y k son los coeficientes del modelo. En este contexto P_w representa una estimación del total de *slices* afectados considerando la propagación de errores entre *slices* de tipo I,P y B.

En el marco del proyecto VQI se decidió generalizar esta idea a distintos tipos de GOP, particularmente se propuso por el grupo de investigación un algoritmo que detecta cuales *slices* fueron afectados debido a la pérdida de paquetes TS y luego dependiendo del tipo de *slice* I, P o B propaga el error dependiendo del GOP de manera automática. Para lograr este objetivo se modificó el software *FFmpeg* para que arroje información sobre los *slices* y *cuadros* que tuvieron errores, luego de una recorrida por el archivo generado se arma la estructura utilizada de *slices* por *cuadro* en el clip de video. Luego se identifica cuales se decodificaron con errores por pérdida de paquetes TS. En el siguiente paso se hace una recorrida y se propagan los errores dependiendo del tipo de *slice* I, P o B. Si se trata de un I se afecta hasta el siguiente I, si se trata de un P se afecta hasta el siguiente I y los B hacia atrás y por último si se trata de un B el error no se propaga. De esta manera se obtiene una estimación del porcentaje de *slices* afectados en el total de *slices* del clip de video, así como el porcentaje de cuadros afectados en el total de cuadros del video. En caso de tratarse de un *slice* por cuadro, como es el caso de los videos codificados de los conjuntos Set I HD y Set I SD, el resultado será el mismo. En cambio en el caso de los conjuntos *Set II HD* y *Set II SD* se tienen clips codificados hasta con 6 *slices* por cuadro, en este caso la relación entre porcentaje de cuadros y *slices* no tiene porque mantenerse.

Con este nuevo enfoque se propone calcular P_w de acuerdo a la ecuación 17:

$$P_w = \frac{C_A}{C_T} \tag{17}$$

P_w en este caso es el porcentaje entre el número de cuadros afectados (C_A) obtenido tras utilizar el algoritmo de propagación en el GOP sobre el total de cuadros (C_T) en la secuencia. Con el objetivo de obtener una predicción de Ip (Ip_{Pred}) se evaluaron dos curvas de ajuste presentadas en las ecuaciones 18 y 19.

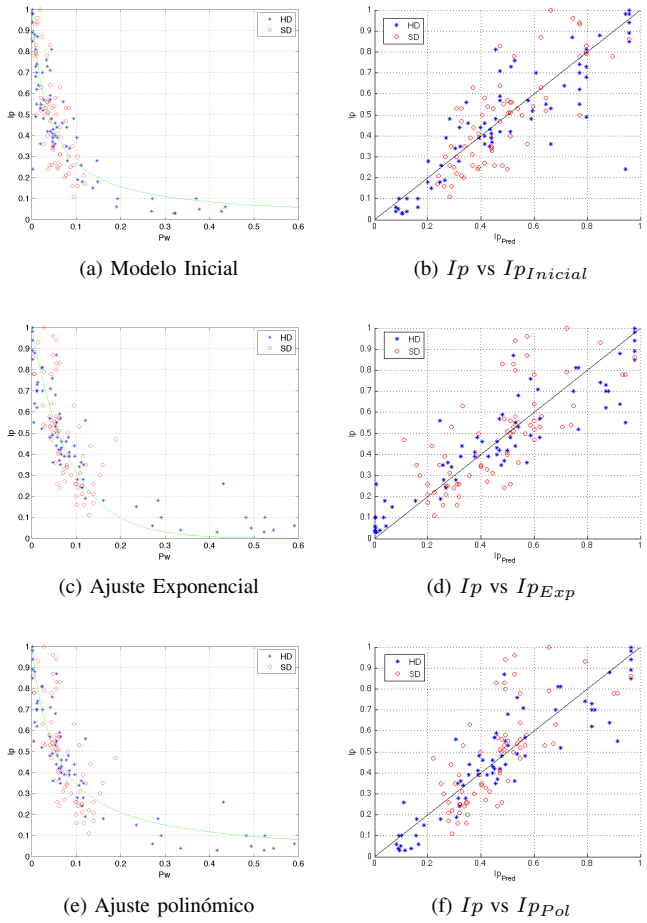
$$Ip_{Exp} = e^{-\alpha P_w} \tag{18}$$

$$Ip_{Pol} = \frac{1}{1 + \beta P_w} \tag{19}$$

Donde α y β son los parámetros de calibración, respectivamente.

Con el fin de calibrar y medir el desempeño de los modelos propuestos se utilizó aquellos clips de video del conjunto de entrenamiento *Set I* con pérdida de paquetes ($Ip < 1$). En la figura 5 (a) se puede observar la predicción $Ip_{Inicial}$ en función de P_w (ecuaciones 15 y 16). Mientras que en la figura 5 (b) se observa la dispersión entre Ip y su predicción $Ip_{Inicial}$, en la figura 5 (c) y (e) se muestran los resultados obtenidos tras aplicar ambas curvas de ajuste (ecuaciones 18 y 19), respectivamente, al modelo de P_w de la ecuación 17. Mientras que en las figuras 5 (c) y (f) se tiene la dispersión entre Ip y sus predicciones Ip_{Exp} e Ip_{Exp}, respectivamente.

En la tabla VI se presentan la PLCC y el RMSE obtenido para cada uno de los modelos de estimación de Ip propuestos incluyendo el modelo Inicial. Se puede observar que los modelos de estimación de Ip utilizados tienen un desempeño similar sobre el conjunto de datos seleccionado. Sin embargo, cabe destacar que los modelos Ip_{Exp} e Ip_{Pol} extienden el uso



(a) Modelo Inicial (b) Ip vs $Ip_{Inicial}$

(c) Ajuste Exponencial (d) Ip vs Ip_{Exp}

(e) Ajuste polinómico (f) Ip vs Ip_{Pol}

Figura 5: Ilustración de los resultados en la estimación de Ip, en (a) se puede observar el $Ip_{Inicial}$ en función de P_w y en (b) la dispersión entre Ip y su predicción $Ip_{Inicial}$, en (c) se presenta el ajuste exponencial, en (d) se tiene la dispersión de Ip en función de su predicción Ip_{Exp}, en (e) se presenta el ajuste polinómico y en (f) se tiene la dispersión de Ip en función de su predicción Ip_{Pol}

Cuadro VI: Comparación de los resultados obtenidos para los modelos de estimación de Ip propuestos

Modelo	PLCC	RMSE
$Ip_{Inicial}$	0,837	0,141
Ip_{Exp}	0,844	0,146
Ip_{Pol}	0,832	0,143

del modelo Inicial a otros tipos de GOP, no solo al utilizado para la codificación de las secuencias seleccionadas.

IV-C. MOSp - Predicción del MOS

Una vez calibrados los coeficientes de los modelos asociados a Ic e Ip se puede obtener una estimación del MOS para una secuencia dada tras aplicar la ecuación general 9. En la tabla VII se compara el desempeño de los modelos propuestos al ser aplicados en el conjunto *Set I HD* & *Set I SD* y luego en el *Set II HD* & *Set II SD*. El modelo $MOSp_{Inicial}$ representa el modelo paramétrico inicial propuesto en [47], los valores de PLCC y RMSE presentados en esta tabla fueron extraídos de [47]. Por otro lado los modelos $MOSp_{Exp}$ y $MOSp_{Pol}$ utilizan el mismo modelo para Ic y distintos modelos de Ip. Por un lado el modelo $MOSp_{Exp}$ utiliza las ecuaciones 18 y 17 para realizar la estimación de Ip, mientras que el modelo $MOSp_{Pol}$ utiliza las ecuaciones 19 y 17. Como medida de desempeño se utilizó la Correlación de Pearson (PLCC) y la raíz del error cuadrático medio (RMSE) entre la predicción del MOS y el MOS subjetivo obtenido en cada conjunto de videos donde estos modelos fueron aplicados.

Cuadro VII: Comparación de los resultados obtenidos en entrenamiento (Set I HD & SD) y validación (Set II HD & SD) de los modelos de estimación del MOS propuestos.

Modelo	Set I HD & SD		Set II HD & SD	
	PLCC	RMSE	PLCC	RMSE
$MOSp_{Inicial}$	0,91	0,42	0,81	0,80
$MOSp_{Exp}$	0,9031	0,434	0,9122	0,55
$MOSp_{Pol}$	0,8989	0,422	0,9030	0,50

En la tabla VII se puede observar que todos los modelos propuestos tienen un desempeño similar en el *Set I*. Por otro lado en el *Set II* los modelos $MOSp_{Exp}$ y $MOSp_{Pol}$ superan en desempeño al modelo $MOSp_{Inicial}$. Esta diferencia es esperable dado que al ser un conjunto de videos grabados de aire contienen distintas estructuras de GOP y distinto número de *slices* por cuadro según la estación de TV utilizada para registrar el contenido (*transport stream* recibido). Por otro lado es alentador el rendimiento de los modelos $MOSp_{Pol}$ y $MOSp_{Exp}$ en el *Set II*.

Se observa por medio de los resultados obtenidos, que los modelos $MOSp_{Exp}$ y $MOSp_{Pol}$ superan en desempeño al modelo $MOSp_{Inicial}$, comprobándose además su utilización con señales reales.

V. SMTVD (Sistema de Monitorización de la señal de TV Digital)

Aunque en los servicios de comunicaciones es común medir la calidad del servicio prestado a los usuarios finales en general este no es el caso de la Televisión Abierta. Si bien la calidad del contenido se controla mientras se produce, codifica, distribuye y finalmente transmite, una vez que la señal se emite por la antena no es habitual que se supervise de forma permanente. Los reguladores normalmente actúan de manera reactiva, cuando los usuarios se quejan del servicio o un radiodifusor realiza una denuncia, por ejemplo a causa de interferencia. Es por este motivo que en el marco del proyecto SMTVD, se propone un sistema de monitorización permanente y automático de la señal de TV Digital ISDB-T.

Este sistema se diseñó para medir un amplio rango de factores que afectan a la señal, desde parámetros físicos como potencia recibida, indicadores de artefactos (efecto de bloques, borrosidad, etc), hasta el resultado final, que es la calidad de experiencia por parte de los espectadores. Dado que la *QoE* involucra múltiples factores que afectan la percepción global, se acotó el problema haciendo énfasis en la estimación de la calidad percibida de video y audio. La finalidad del sistema es apoyar la regulación con un enfoque complementario, por

ejemplo frente a denuncias, para mejorar la calidad de experiencia y posiblemente contar con un conocimiento a priori en la predicción de la opinión de los televidentes. Este sistema no solo sirve a los radiodifusores, sino a los reguladores y agencia de protección al consumidor. Por ello es independiente de los radiodifusores dado que no requiere información adicional de las estaciones de TV que son monitorizadas. En particular no se emplea la señal original de estudio.

En la figura 6 se puede observar una ilustración a alto nivel de la arquitectura del sistema SMTVD.

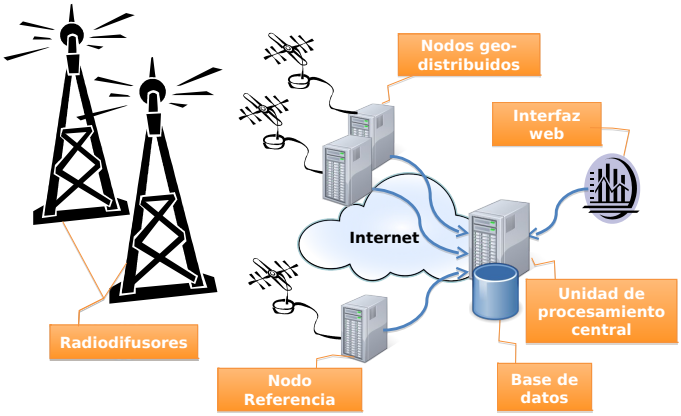


Figura 6: Ilustración a alto nivel de la arquitectura del sistema de monitorización en tiempo real de la señal de TVD (SMTVD)

El sistema de monitorización consiste de nodos remotos geográficamente distribuidos, un nodo de referencia y una unidad de procesamiento central con una interfaz web para la visualización de los datos. Los nodos remotos están conformados por sintonizadores de TV Digital de tipo *dongle*, que se conectan por medio de un puerto USB a dispositivos de procesamiento de bajo costo. Estas estaciones pueden ser ubicadas en los distintos sitios donde se desee monitorear la cobertura digital. La principal función de estos nodos es sintonizar la señal de TVD en un canal físico predefinido, extraer parámetros físicos (como la potencia de la señal) así como del bitstream (por ejemplo, detección de pérdida de paquetes en el *Transport Stream*, (TS)) y de cada una de las señales decodificadas (por ejemplo, afectación de cuadros o slices I, P o B en cada uno de los *elementary streams* (ES) recibidos). Además realizan grabaciones de 10 segundos, donde se registra el TS completo de la señal sintonizada. Si ocurren errores introducidos en el canal de radiofrecuencia que no pueden ser corregidos por los algoritmos de corrección en recepción (por ejemplo, *Reed-Solomon*) los TS registrados podrán contener pérdida de paquetes TS.

Con el objetivo de medir independientemente artefactos de codificación respecto de aquellos introducidos por errores en el canal, el sistema cuenta con un nodo especial, denominado como nodo de referencia. Este nodo debe ser ubicado en un sitio privilegiado (preferentemente con línea de vista con la antena emisora y una antena de alta ganancia) donde se obtenga una óptima recepción, siendo la señal digital sintonizada en este sitio idéntica a la señal emitida por el radiodifusor. Si bien en un sitio fijo no se puede asegurar una recepción continua sin errores, en un sistema distribuido pueden encontrarse puntos donde la probabilidad de error sea prácticamente nula. Se puede imaginar un escenario alternativo, si los radiodifusores entregaran su señal por fibra óptica u otro canal sin pérdidas el nodo de referencia se podría alimentar de esta señal. Si bien esta aproximación es técnicamente buena, como se mencionó previamente, se prefiere que la implementación sea independiente de los radiodifusores por razones prácticas y de regulación. De todos modos, en la implementación del sistema se previó el uso de edificios de la Universidad que pueden lograr este requisito de recepción. Por otro lado en caso de ser necesario se prevé el uso de varios nodos de referencias, uno por señal.

Todos los nodos (remotos y de referencia) se comunican con la unidad de procesamiento central a través de Internet utilizando mecanismos de comunicación seguros. Periódicamente o por demanda, los nodos envían a la unidad de procesamiento central las distintas medidas tomadas y el porcentaje de pérdida de paquetes detectado en las grabaciones realizadas. Toda la información recolectada por la unidad de procesamiento central es guardada en una base de datos. Esta unidad utiliza tecnología CEP (Complex Event Processing) [49] para detectar situaciones inusuales en el servicio. Cuando un evento de pérdida de paquetes es recibido desde un nodo remoto, el sistema dispara un proceso que solicita al nodo remoto y al nodo referencia las correspondientes grabaciones (TS). Cabe resaltar que los TS enviados incluyen cuadros anteriores y posteriores a la situación inusual permitiendo un análisis completo.

En la siguiente sección se describe la evaluación de calidad de video realizada en el sistema SMTVD.

V-A. *Evaluación de calidad de video en el sistema SMTVD*

Con el objetivo de realizar una evaluación global y objetiva de la calidad de video en cada uno de los nodos del sistema se utiliza la ecuación general 9, utilizada por el modelo paramétrico descrito en la sección IV. Este modelo general parte de la existencia de dos componentes abreviadas, Ic e Ip, donde, Ic representa la calidad dada por el material en sí mismo, y el proceso de codificación e Ip representa la degradación de la calidad de la señal debido errores en transmisión. En tanto, bajo las hipótesis y consideraciones tomadas para el sistema la componente de calidad Ic debe ser medida en el nodo de referencia. En cambio, la componente Ip puede ser calculada únicamente con la señal obtenida en los nodos remotos donde hayan existido posibles errores en transmisión (por medio de un modelo NR), o puede obtenerse realizando una comparación entre la señal recibida en el nodo de referencia y la señal recibida en el nodo remoto (por medio de un modelo FR o RR).

A continuación se describe el cálculo de las componentes Ic e Ip en el sistema SMTVD.

V-B. Cálculo de Ic

Con el fin de realizar el cómputo de Ic en el sistema SMTVD se utiliza el modelo paramétrico de la ecuación 11 presentado en la sección IV. Este modelo se aplica a los ES de video asociados al TS recibido y grabado para los canales de TV predefinidos en el nodo de referencia. Cada uno de los TS grabados tiene una duración de 10 segundos. Como ha sido mencionado previamente el modelo tiene como insumo el bitrate, la resolución y el SAD de cada una de las programaciones emitidas. El bitrate y la resolución se extraen del bitstream utilizando la aplicación *FFmpeg* [8]. Mientras que para extraer el SAD se utilizó algunas de las herramientas presentadas en [50], puntualmente el *plugin* denominado como *MVtools_v2* con *AVISynth*[9] y *VirtualDub* [10]. Específicamente se utilizó la función "MAnalyse". Estas aplicaciones son invocadas de manera automática en el nodo referencia. Con este procedimiento se obtiene el Ic correspondiente para cada una de las programaciones emitidas por canal de TV. Posteriormente este valor se envía a la unidad de procesamiento central para el posterior cómputo de calidad global ($MOSp$) según la ecuación 9.

En la tabla VIII se muestra el costo en términos de tiempo de ejecución de la implementación del modelo de estimación de Ic seleccionado y analizado en este trabajo. Cabe mencionar que la integración e implementación del modelo seleccionado de estimación de Ic al sistema de monitorización fue realizada en el marco del proyecto SMTVD por un subgrupo de trabajo del proyecto. El hardware utilizado fue un procesador Intel Core i7-4702HQ con 16GB RAM con un sistema operativo Ubuntu (Linux). Con el fin de obtener las medidas de tiempo se utilizó una grabación de 10 segundos realizada por el sistema en resoluciones HD y otra en resolución SD.

Cuadro VIII: Tiempo de ejecución del modelo asociado a Ic

Indicador	HD	SD
Ic	13.3 seg	8.1 seg

Como puede verse Ic puede calcularse de manera continua y en tiempo real para señales en resolución SD. Para señales HD, se podría tener un valor de Ic cada 15 segundos, cuatro veces por minuto, realizando solo la estimación sobre 10 de los 15 segundos. La mayoría del tiempo para el procesamiento de Ic se consume en el cálculo del SAD, al utilizar la aplicación *virtualdub*. A futuro, este módulo se puede mejorar realizando el cálculo en un lenguaje con mejor desempeño en términos de tiempo de ejecución, haciendo posible el cómputo de Ic de manera continua y en tiempo real también para señales en resolución HD.

V-C. Cálculo de Ip

En esta sección se muestra como al utilizar modelos FR se mejora el desempeño de los modelos paramétricos NR de estimación de Ip previamente propuestos. Los resultados obtenidos en esta sección fueron publicados en [51], [52]. Acorde a las ecuaciones 17, 18 y 19 la degradación de calidad debido a errores en transmisión depende del parámetro Pw, el cual es una estimación del total de cuadros afectados considerando la propagación de errores en el GOP cuando se ven afectados *slices* I o P. Teniendo en cuenta esta observación, Pw también puede ser expresado según la ecuación 20:

$$Pw = 1 - \frac{1}{N}\sum_{i=1}^{N} Q(i) \tag{20}$$

Donde N es el número total de cuadros del clip de video, $Q(i) = 0$ si el cuadro i se encuentra afectado (directa o indirectamente) y $Q(i) = 1$ si el cuadro i no se encuentra afectado. Con esta aproximación la degradación perceptual de calidad se realiza mediante una clasificación binaria entre los cuadros afectados y no afectados. En este contexto solo se toma en cuenta el porcentaje de cuadros afectados sin importar su grado de distorsión. Sin embrago bajo las hipótesis tomadas en el sistema, es posible contar con la señal sin degradaciones debidas a errores introducidos en el canal de transmisión. En tanto, por medio de modelos objetivos (FR o RR) es posible cuantificar dicha degradación de calidad percibida entre ambos clips, inclusive cuadro a cuadro, para los clips de video grabados en el nodo remoto y el nodo de referencia. En consecuencia es razonable pensar que de esta manera se logre mejorar la estimación del grado de distorsión Ip entre la señal recibida respecto de la señal emitida.

Existen distintos modelos FR/RR de estimación de calidad de video (VQA) así como de imágenes (IQA) que pueden ser utilizados para esta tarea [12]. En este trabajo se evaluaron los siguientes modelos, previamente descritos en la sección II: PSNR, SSIM, MS-SSIM, VIFp, PSNR-HVS, VQM *General* y VQM *LowBw*.

Con el fin de obtener la degradación cuadro a cuadro entre la señal de referencia y la degradada los modelos SSIM, MS-SSIM y VIFp parecen ser buenos candidatos, dado que al aplicarlos arrojan un valor de 0 cuando el cuadro degradado se encuentra totalmente distorsionado y 1 cuando los cuadros son idénticos. Con el objetivo de obtener un valor global al utilizar estos modelos para toda la secuencia de cuadros se puede utilizar la ecuación 21.

$$P_w = 1 - \frac{1}{N}\sum_{j=1}^{N} Q_{v_x}(i) \tag{21}$$

Donde, N es la cantidad de cuadros, $Q_{v_x}(i)$ es la evaluación de calidad de imagen del modelo x (SSIM, MS-SSIM o VIFp) para el cuadro i de la secuencia degradada y original. Comparando las ecuaciones 20 con 21 uno puede notar que en el primer caso los cuadros afectados cuentan como 0, sin importar el grado de degradación, mientras que en el segundo caso cada cuadro afectado tiene su propio peso entre 0 y 1 dependiendo del grado de distorsión estimado por cada modelo.

Para ilustrar este punto con un ejemplo en la figura 7 se observa el resultado de los modelos SSIM, MS- SSIM y VIFp para cada uno de los cuadros sobre 4 clips de videos

[8] *FFmpeg*, en línea , https://www.ffmpeg.org
[9] En línea, http://avisynth.nl/index.php/Main_Page
[10] En línea, http://www.virtualdub.org/

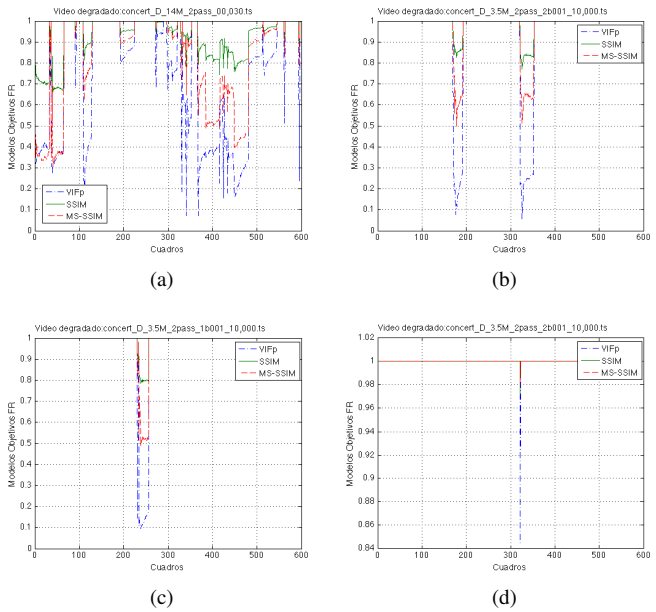


Figura 7: Resultado de los modelos VIFp, SSIM y MS-SSIM sobre cada uno de los 600 cuadros que conforman un clip de video de la base *Set I HD* con (a) distribución uniforme de pérdida de paquetes (b) dos ráfagas (*bursts*) de errores (c) una ráfaga de errores. Por último en (d) se observa un ejemplo donde solo se vio afectado un cuadro, en este caso se afectó un *slice* de tipo B

tomados del *Set I HD*. En las abscisas se representa cada uno de los cuadros del video correspondiente mientras que en las ordenadas se representa el resultado de cada modelo al realizar una comparación cuadro a cuadro entre el original y el degradado. En la gráfica (a) se utilizó un clip con una distribución de pérdida de paquetes uniforme, en (b) un clip con dos ráfagas (*bursts*) de errores , en (c) un clip con una ráfaga de errores y por último en (d) se observa un ejemplo de un clip con sólo una ráfaga de errores donde sólo se vio afectado un cuadro, en este caso debido a un error al decodificar un *slice* de tipo B. Se puede observar que la degradación perceptual de calidad para cada cuadro varía de 0 a 1, siendo 1 cuando es idéntico al cuadro original.

Con el fin de ejecutar los modelos SSIM, MS-SSIM y VIFp, se utilizó la herramienta VQMT disponible en línea http://mmspg.epfl.ch/vqmt.

En adición a estos modelos FR, se evaluó el modelo VQM *General* [31] propuesto por la NTIA y el modelo VQM *LowBw* [35]. Estos modelos proveen directamente para cada par de videos (original y degradado) un valor entre 0 cuando no hay diferencias perceptible y 1 cuando hay máxima distorsión. En este caso P_w se puede asociar directamente con este valor. Con el fin de aplicar estos modelos se utilizó la implementación de referencia provista por NTIA en línea (http://www.its.bldrdoc.gov/resources/video-quality-research/software.aspx).

Aunque hoy en día no sea una medida aceptada de calidad percibida por observadores humanos [20] igualmente se decidió evaluar el PSNR por ser una medida históricamente utilizada. Recordando la formulación del PSNR para su cálculo entre dos imágenes, ecuación 2, se debe notar que cuando ambas imágenes son idénticas el resultado es ∞. En tanto, cabe destacar que si se desea aplicar el PSNR cuadro a cuadro entre la señal recibida en el nodo remoto y la señal recibida en el nodo de referencia, en el caso que no haya errores en transmisión las señales digitales serán idénticas, por lo tanto también los cuadros de ambos clips. En el caso que haya pérdida de paquetes, éstas afectarán algunos de los cuadros debido a la propagación de errores en el GOP, mientras el resto de los cuadros permanecerán idénticos. Esto impide realizar un promedio del PSNR en el total de cuadros del clip de video dado que para cada uno de los cuadros sin afectación dará ∞. Con el objetivo de superar este problema y como es sugerido en [53] para transmisiones de video sobre canales con pérdida, el PSNR sobre el video se calcula basándose en la media aritmética (Log-Av) en vez de la media geométrica (Av-Log). Esto es, en vez de calcular el PSNR de acuerdo a la ecuación 2 para cada par de cuadros y luego realizar un promedio entre la cantidad total de medidas obtenidas, se calcula el MSE, según la ecuación 1 y luego se realiza un promedio sobre todas las medidas obtenidas cuadro a cuadro, obteniendo un MSE final único. Luego se calcula el PSNR para esta medida obtenida de MSE según la ecuación 2. Con esta aproximación se logra que el PSNR sea ∞ solo cuando los videos son idénticos, en cuyo caso $Ip = 1$. Por último, en adición al PSNR te utilizó el método PSNR-HVS como modelo objetivo de calidad. Para tal fin se utilizó la implementación provista de este modelo en el software VQMT.

Cabe destacar que todos estos modelos en la literatura han sido verificados comparando imágenes o videos con su respectiva fuente original, de alta calidad y sin degradación. En el contexto donde se quieren utilizar si bien se cuenta con una señal de referencia sin degradación debido a errores en el canal de transmisión no se puede garantizar que no existan degradaciones debidas a compresión o problemas en la adquisición de la imagen. De hecho, distintos radiodifusores pueden utilizar distinto bitrate en la compresión H.264, introduciendo distintos grados de distorsión. Por lo tanto estos modelos se deben validar en las condiciones donde serán aplicados por el sistema SMTVD. Con el fin de comparar el desempeño de estos modelos objetivos en la estimación de Ip se utilizó el conjunto de clips de video *Set I HD y Set I SD*. Aquí se seleccionaron aquellos clips con pérdida de paquetes TS ($Ip < 1$) con su correspondiente clip de referencia (con igual esquema de codificación (*bitrate*) y sin pérdida de paquetes TS). Además, teniendo en cuenta que el objetivo principal del sistema es obtener medidas en tiempo real uno de los aspectos que se evaluó fue el costo en términos de tiempo de ejecución consumido por cada modelo sumado a su desempeño en términos de correlación con el resultado subjetivo.

En general para realizar una comparación de desempeño entre modelos objetivos se utiliza un ajuste no lineal [54] para convertir el resultado de los modelos a una escala lineal específica respecto del resultado subjetivo (Ip).

En este trabajo se utilizó la transformación no lineal pre-

sentada en 22:

$$Ip_{pred} = t_1 \left(0,5 - \frac{1}{1 + e^{(t_2(x - t_3))}} \right) + t_4 \qquad (22)$$

Donde Ip_{pred} es la predicción de Ip y x es el resultado de los modelos SSIM, MS-SSIM, VIFp, VQM *General*, VQM *LowBw*, PSNR y PSNR-HVS, respectivamente. Por otro lado t_1, t_2, t_3, t_4 son los coeficientes de calibración de la regresión.

En la figura 8 se muestra la dispersión entre el valor de Ip y su predicción tras aplicar la regresión de la ecuación 22 a cada uno de los modelos objetivos utilizados (x). Con el símbolo "o" se identifican los videos en resolución SD y con "$*$" los videos en resolución HD.

En la tabla IX se muestra el desempeño de los modelos propuestos, se utilizó la correlación de Pearson y el RMSE entre los valores de Ip_{Pred} e Ip y las correlaciones de Spearman y Kendall entre el resultado x asociado a cada modelo y los valores de Ip. Además se tomó el tiempo de ejecución de cada modelo teniendo en consideración el proceso de decodificación y el proceso de alineación temporal (que se describe en la siguiente sección) entre las secuencias de referencia y degradada. Para ello se utilizó un ordenador con CPU Intel Core i7-4770 @ 3.40GHz con 32 GB RAM y secuencias en resolución HD y SD de 10 segundos de duración.

En comparación con los modelos paramétricos NR inicialmente propuestos, cuyo desempeño se muestra en la tabla VI, se puede observar que los modelos Ip_{PSNR}, $Ip_{VQMLowbw}$,$Ip_{MS-SSIM}$, Ip_{SSIM} e Ip_{VIFp} los superan en desempeño en términos de alta correlación de Pearson (PLCC) y menor error cuadrático medio (RMSE). Con el objetivo de realizar un análisis estadístico del desempeño de los modelos se siguen los criterios adoptados por VQEG en [34] utilizando la Transformación z de Fisher. El test estadístico asume la hipótesis nula (H0) de que no hay diferencia significativa entre los coeficientes de correlación de Pearson (PLCC). En este caso los modelos $Ip_{VQMLowbw}$, Ip_{SSIM} e Ip_{VIFp} muestran una mejora estadística respecto al modelo $Ip_{Inicial}$ con una nivel del confianza del 97 %. Este tipo de resultado concuerda con lo esperado, dado que el modelo paramétrico inicial es un modelo NR mientras que estos son modelos FR.

Por otro lado, cabe destacar que los modelos paramétricos propuestos tienen un tiempo de ejecución menor con respecto a los modelos FR/RR propuestos tomando menos de 6 segundos para realizar la estimación de Ip para un clip en HD y menos de 2 segundos para un clip en SD (ambos de 10 segundos). Entre los modelos FR/RR con mejor desempeño en términos de PLCC y RMSE el modelo Ip_{SSIM} es el que permite, sobre las señales adquiridas (degradada y de referencia), correr en menos tiempo tomando 8 segundos para un clip en SD y 33 segundos para un clip en HD. Si bien es una desventaja que para clips de 10 segundos en HD no se pueda realizar una medida en tiempo real de manera continua, dos factores hacen que la utilización de este modelo sea aceptable. En primer lugar los modelos de estimación de Ip no deben correr todo el tiempo, solo cuando hay pérdida de paquetes en alguno de los nodos remotos, por otro lado el modelo FR corre en un solo lugar permitiendo actualizar el modelo y el hardware de manera independiente. De ser necesario se puede introducir más poder de cómputo por medio de mejoras en el equipamiento para reducir los tiempos de procesamiento dentro del sistema de monitorización.

V-D. MOSp - Predicción del MOS

Con el fin de realizar la estimación del MOS empleando los distintos modelos propuestos se utiliza la formula global 9, con el componente Ic dado por la ecuación 11 y el componente Ip correspondiente a los distintos modelos de estimación de Ip. La nomenclatura utilizada es $MOSp_x$ donde x refiere a cada uno de los distintos modelos objetivos FR/RR/NR empleados para realizar la estimación de Ip. Se comienza el análisis al aplicar los distintos modelos $MOSp_x$ en la misma selección de clips donde fueron verificados los modelos FR para la estimación de Ip. En la tabla X se muestra el desempeño en términos de PLCC y RMSE obtenido en cada caso.

Cuadro X: Desempeño de los modelos FR/RR/NR propuestos sobre los clips de video con pérdida de paquetes TS ($Ip < 1$) de los conjuntos *Set I HD* y *Set I SD*, se utiliza la PLCC (Correlación de Pearson) y RMSE (Raíz del Error Cuadrático Medio).

Modelo	PC	RMSE
$MOSp_{PSNR}$	0.858	0.394
$MOSp_{VQMGeneral}$	0.810	0.445
$MOSp_{VQM_{Lowbw}}$	0.864	0.386
$MOSp_{MS-SSIM}$	0.849	0.408
$MOSp_{PSNR-HVS}$	0.775	0.483
$MOSp_{SSIM}$	0.869	0.382
$MOSp_{VIFp}$	0.868	0.383
$MOSp_{VQI_{Inicial}}$	0.816	0.442
$MOSp_{VQI_{Exp}}$	0.834	0.461
$MOSp_{VQI_{Pol}}$	0.818	0.441

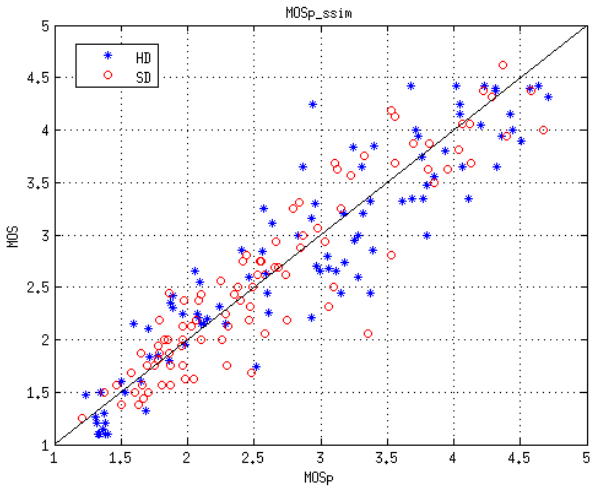
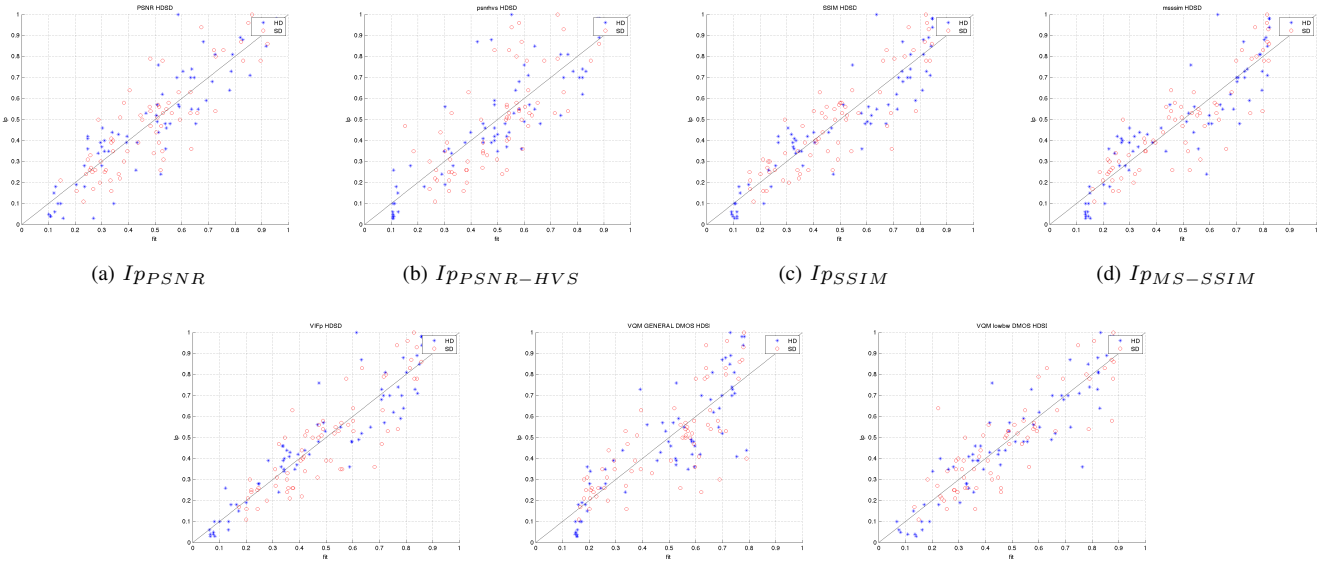


Figura 9: Dispersión $MOS\ vs\ MOSp_{SSIM}$ sobre el conjunto *Set I HD* y *Set I SD*

Como se puede observar los modelos $MOSp_{SSIM}$, $MOSp_{VIFp}$, $MOSp_{VQM_{Lowbw}}$ son los que tienen mejor desempeño (mayor PLCC y menor RMSE). Al realizar nuevamente un análisis estadístico, se tiene que el modelo



(a) Ip_{PSNR} (b) $Ip_{PSNR-HVS}$ (c) Ip_{SSIM} (d) $Ip_{MS-SSIM}$

(e) Ip_{VIFp} (f) $Ip_{VQMGeneral}$ (g) $Ip_{VQMLowBw}$

Figura 8: Desempeño de los modelos Ip_{PSNR}, $Ip_{PSNR-HVS}$, Ip_{SSIM}, $Ip_{MS-SSIM}$, Ip_{VIFp}, $Ip_{VQMGeneral}$ y $Ip_{VQMLowBw}$

Cuadro IX: Desempeño de los modelos FR/RR propuestos sobre los videos con pérdida de paquetes ($Ip < 1$) de los conjuntos *Set I HD* y *Set I SD*, se utiliza la PLCC (Correlación de Pearson), SROCC (Coeficiente de correlación de Spearman), KROCC (Coeficiente de correlación de Kendall) y RMSE (Raíz del Error Cuadrático Medio)

HD & SD	PLCC	SROCC	KROCC	RMSE	Tiempo (HD)	Tiempo (SD)
Ip_{PSNR}	0.88825	0.87154	0.69602	0.11786	18.1 seg	4.3 seg
$Ip_{VQMGeneral}$	0.82847	-0.8392	-0.6679	0.14369	6.9 min	78.7 seg
$Ip_{VQM_{Lowbw}}$	0.90063	-0.89598	-0.73846	0.1115	8.9 min	103.7 seg
$Ip_{MS-SSIM}$	0.89329	0.90466	0.74539	0.11533	40.1 seg	9.0 seg
$Ip_{PSNR-HVS}$	0.80603	0.82695	0.64282	0.15186	44.5 seg	9.9 seg
Ip_{SSIM}	0.90371	0.90793	0.7405	0.10985	33.4 seg	8.0 seg
Ip_{VIFp}	0.90145	0.9062	0.74396	0.11106	53.6 seg	11.3 seg

$MOSp_{SSIM}$ tiene una mejora estadística respecto al modelo $MOSp_{Inicial}$ con un nivel de confianza del 87 %. En base a estos resultados se seleccionó el modelo Ip_{SSIM} como modelo FR para correr en el sistema de monitorización de la señal de TVD, siendo posible actualizar este modelo por ser un módulo independiente en el sistema.

En la figura 9 se muestra la dispersión entre el MOS subjetivo y la predicción del MOS al aplicar el modelo $MOSp_{SSIM}$ sobre todos los clips de video del conjunto *Set I HD* y *Set I SD*. En este caso se obtuvo una PLCC de 0.9227 y un RMSE de 0.3711. Al contrastar estos resultados con los obtenidos por los modelos $MOSp_{Inicial}$, $MOSp_{Exp}$, $MOSp_{Pol}$ en la tabla VII se puede observar que el modelo $MOSp_{SSIM}$ también supera en desempeño a estos modelos paramétricos. Al aplicar el test estadístico en el conjunto total de videos se obtiene que el modelo $MOSp_{SSIM}$ tiene una mejora estadística respecto al modelo inicial con un nivel de confianza del 80 %.

Se observa por medio de los resultados obtenidos que al utilizar modelos FR/RR en lugar de los modelos paramétricos NR propuestos en la estimación de Ip se obtiene un mejor desempeño en la predicción del global del MOS.

V-E. Integración del modelo FR seleccionado al sistema SMTVD

Con el objetivo de aplicar el modelo FR seleccionado en el sistema de monitorización se debe realizar previamente un proceso de alineación temporal entre la secuencia degradada y de referencia, registradas en uno de los nodos remotos y en el nodo de referencia, respectivamente. Si bien en el sistema todos los nodos están sincronizados por medio del protocolo NTP (*Network Time Protocol*), el tiempo exacto en el que se inician las grabaciones puede ser levemente distinto. Por otro lado uno de los factores que pueden afectar la alineación temporal de los clips es la pérdida completa de cuadros debido a pérdida de paquetes, es decir cuando se pierde la información completa de varios cuadros en el clip degradado, en este caso se debe detectar el/los cuadro/s perdido/s y regenerarlo/s, por ejemplo con el cuadro previo, con el fin de mantener

la tasa de cuadros fija. En caso contrario se podría generar una desincronización entre los cuadros del clip degradado y los cuadros del clip de referencia, que afectaría la medida de calidad. Además, resulta adecuado identificar cuadros de inicio y de fin para aplicar el modelo FR cuadro a cuadro entre entre ambas secuencias. El algoritmo de alineación que se propone consta de dos pasos, el primero consiste de la decodificación a formato descomprimido (YUV) de los clips de video contenidos en los *transport streams* registrados por el sistema, en el nodo de referencia y el nodo remoto. El segundo paso consta de un proceso de alineación de cuadros entre las secuencias descomprimidas. A continuación se presenta cada una de las etapas del proceso.

Cada uno de los nodos graba el TS recibido en un momento establecido por el sistema. Cada TS contiene cada uno de los servicios que emite cada canal, por ejemplo un servicio en definición HD, otro servicio en resolución SD y el servicio *one-seg* para receptores móviles. Con el objetivo de obtener cada uno de los flujos elementales (ES) de video en el sistema se utiliza la herramienta *FFmpeg*. Una vez que se obtienen los *ES* se procede a generar las secuencias descomprimidas en formato YUV. Para ello se utiliza el formato de pixel YUV420 y además se fuerza a que se mantenga la tasa de cuadros a la salida, en caso de que se pierda la información de un cuadro completo se repite el cuadro disponible anterior. Por último se genera un archivo que contiene para cada uno de los cuadros decodificados su correspondiente posición en la secuencia y su respectivo código identificador (CRC, *Cyclic Redundancy Check*). La secuencias, de referencia y degradada, de video descomprimidas en formato YUV serán utilizadas por el modelo objetivo FR para efectuar la estimación de calidad. Mientras el archivo que contiene el identificador (CRC) de cada uno de los cuadros decodificados será utilizado para realizar el proceso de alineación de cuadros. Este proceso se describe a continuación.

V-E1. Proceso de alineación de cuadros: Es de esperar que cuando se producen errores en transmisión muchos de los cuadros no sufrirán distorsiones, y por lo tanto los CRC de estos cuadros serán iguales en ambas secuencias (degradada y de referencia). Cuando un cuadro decodificado en la secuencia degradada se encuentra parcialmente afectado este será diferente a su par en la secuencia de referencia, dado que la información para decodificar correctamente el cuadro completo está perdida (por ejemplo, Vectores de Movimiento y/o Macro Bloques perdidos). En este escenario el CRC de los cuadros afectados de la secuencia degradada será distintos al de sus pares en la secuencia de referencia. Por otro lado, es posible que se pierda por completo la información de un cuadro entero, incluso de varios cuadros enteros sucesivos. En este caso, como se mencionó anteriormente, se configuró el *FFmpeg* para repetir el último cuadro disponible tantas veces como sea necesario, para mantener la tasa de cuadros y no generar una discontinuidad en la secuencia. Con el fin de ilustrar el proceso de alineación de cuadros, en la tabla XI se presenta un ejemplo sencillo con dos secuencias, una de referencia (obtenida en el nodo de referencia) y otra degradada (obtenida en un nodo remoto) con pérdida de paquetes, cada una de 13 cuadros, donde se presenta la información obtenida de los CRC para cada uno de los cuadros decodificados para cada una de las secuencias y en la primer columna el número de cuadro.

Cuadro XI: Proceso temporal de alineación de cuadros

Cuadro		CRC Cuadro Referencia		CRC Cuadro Degradado
0		70c262849b29003bb98c39cc13b4607f		1ee4063e604f2366d1b9297a61f9150c
1		e3f242ab7c5f184e9375eb22c06ad136	†	9de5a9194ac4bc03f4b43689f6506bc7
2	†	9de5a9194ac4bc03f4b43689f6506bc7		df49c93d3fcbfc4f822c97d30e13925c
3		df49c93d3fcbfc4f822c97d30e13925c		b59cd299a0de8043971da3daac0b15e0
4		b59cd299a0de8043971da3daac0b15e0	‡	3e60f5f86408000823b5683cadc4f311
5	‡	81bd8ff877f4350cc9efcb1590e438ad		488640803e60f0b5683ca0823dc4f311
6		488640803e60f0b5683ca0823dc4f311		1ea3b397d3cad00cdbf8304afa7ec007
7		1ea3b397d3cad00cdbf8304afa7ec007	♣	1ea3b397d3cad00cdbf8304afa7ec007
8		2bb1d7aeacc9d3eaac236334ba7c5169	♣	1ea3b397d3cad00cdbf8304afa7ec007
9		23a0399c43668daf9c02c9fa3b65fc56		ebac9ea4fc955aa0b8f041eb042e6349
10		ebac9ea4fc955aa0b8f041eb042e6349		3e9d36db20ae65bf9773023f64b40328
11		3e9d36db20ae65bf9773023f64b40328	⋆	5849681fd4bf8d654ec0cffbed9d2fe2
12	⋆	5849681fd4bf8d654ec0cffbed9d2fe2		fd6efef78c2a49418acb4089efddbd0d

Se puede observar que en la secuencia degradada los cuadros identificados con ♣ tienen el mismo CRC, que a su vez es igual al CRC del cuadro número 6. Este es un ejemplo donde se perdieron dos cuadros completos sucesivos debido a la pérdida de paquetes y como resultado se copió el último cuadro disponible (6). El proceso de alineación de cuadros tiene dos salidas, que son el sesgo u *offset* (en unidad de cuadros) que representa el número de posición del cuadro inicial en las secuencia degradada y de referencia. Es a partir de estos cuadros que los modelos FR comenzarán a realizar la comparación. Con este fin se forma un vector de comparación temporal T de largo N, en este caso $N = 13$. Luego se procede a realizar una comparación entre los identificadores CRC entre los cuadros de ambas secuencias. En caso de que sean iguales en la correspondiente entrada del vector T se fijará un valor de 1, en caso contrario de 0. Se realiza esta comparación a lo largo todos los cuadros y luego se suma el resultado de cada entrada del vector T y se guarda su valor. Se realiza este proceso reiteradamente corriendo un número de veces arbitrario hacia arriba y abajo los valores de CRC de la columna de la secuencia degradada respecto de la original y se realiza el mismo proceso tomando nota del corrimiento asociado. Una vez finalizado el proceso de correlación se encuentra el máximo valor obtenido y su correspondiente corrimiento. La idea es encontrar el corrimiento donde la correlación se maximiza. En el caso del ejemplo el cuadro inicial en ambas secuencias se encuentra identificado con † , posición número 2 en la secuencia original y en la posición número 1 en la secuencia degradada. En este caso se genera un *offset* igual a dos cuadros en la secuencia original y de un cuadro en la secuencia degradada. Una vez que los cuadros se encuentran alineados queda claro que el cuadro decodificado de la secuencia degradada identificado con ‡ se vio parcialmente afectado dado que el anterior y su par en la secuencia de referencia tienen un CRC distinto.

Una vez que se realiza el proceso de alineación temporal se pueden aplicar el modelo FR seleccionado, para ello se utiliza el software VQMT. Éste se debió adaptar para además de ingresar la secuencia original y degradada, incluya estos *offsets*. De esta manera internamente se aplica el modelo FR seleccionado con los cuadros perfectamente alineados.

VI. Conclusión

Esta tesis contiene aportes específicos en el área de evaluación objetiva de calidad de video con aplicación en Televisión Digital Abierta (TVD). Particularmente se propone y desarrolla un novedoso método que combina modelos objetivos con referencia completa (FR) y sin referencia (NR) para efectuar una medición global de la calidad percibida de video en la señal de TVD. Esta aproximación se basa en la utilización de técnicas modernas de evaluación de calidad de video sobre un sistema de monitorización basado en nodos receptores de TVD distribuidos geográficamente en el área de cobertura de una estación de TV. Este sistema cuenta con un nodo de referencia bajo la hipótesis de que es capaz de registrar las señales digitales emitidas sin distorsiones debido a errores en el canal, es decir pérdida de paquetes TS (*Transport Stream Packets*). Por otro lado cuenta con nodos remotos distribuidos, que permiten adquirir las señales emitidas al aire en distintas ubicaciones. Estas señales sí pueden verse afectadas por pérdida de información debido a errores en el canal de transmisión.

Se muestra que al utilizar modelos objetivos de estimación de calidad de video de tipo FR que comparen ambas señales, de referencia y degradada, adquiridas en el nodo de referencia y remoto, se obtiene una mejora estadística significativa en la estimación de la degradación perceptual de calidad debido a la pérdida de paquetes TS en comparación con los modelos paramétricos NR inicialmente propuestos.

Particularmente el modelo FR seleccionado y utilizado por el sistema para estimar la degradación de calidad debido a pérdida de paquetes es el modelo SSIM, logrando una mejora estadística respecto del modelo paramétrico NR inicialmente propuesto con un nivel de confianza del 97 %. Si bien en los resultados preliminares su aplicación permitiría su ejecución en clips de video en resolución SD en tiempo real, no es el caso para los clips de video en resolución HD. Igualmente los tiempos obtenidos son aceptables para el funcionamiento del sistema, siendo posible por ser un módulo independiente en caso de ser necesario aumentar la capacidades de cómputo, disminuyendo de esta manera los tiempos de ejecución.

Con el fin de tomar una medición global y no solo la degradación de la calidad percibida de la señal debido a la pérdida de paquetes TS en los distintos nodos del sistema, se realiza una medida de calidad sobre la señal digital de video registrada en el nodo de referencia, sin errores debido a pérdida de paquetes (solo con degradaciones debidas a artefactos de codificación y adquisición de la imagen). Para ello se analizan las degradaciones asociadas al proceso de compresión de video (compresión H.264/AVC) y se utiliza un modelo paramétrico NR que toma características del contenido y del flujo de bits de información de la señal de TVD, particularmente la resolución, el *bitrate* y un indicador de complejidad de codifiación del contenido del clip de video, denominado SAD. Con estas dos medidas el sistema es capaz de realizar una estimación global de la calidad en cada nodo de medición.

Como trabajo a futuro con el fin de extender la verificación y validación sobre distintas secuencias de video es de interés realizar nuevas tandas de evaluaciones subjetivas sobre secuencias (degradadas y de referencia) con errores debido a pérdida de información en el canal de transmisión registradas por el sistema de monitorización. Por otro lado con el fin de extender el estudio de medición de calidad percibida de video a señales audiovisuales, se prevé la inclusión de calidad de audio con el fin de lograr modelos que estimen la calidad multimedia (audiovisual) del contenido transmitido. Si bien en este trabajo el foco de aplicación de los modelos de calidad ha sido en TVD, también resulta de interés extender el estudio de calidad percibida y su aplicación a nuevas o futuras normas de TV, por ejemplo TV Híbrida, IPTV y nuevas resoluciones de pantalla (4K).

Referencias

[1] P. Le Callet, S. Möller, A. Perkis *et al.*, "Qualinet white paper on definitions of quality of experience," *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)*, Mar. 2012.

[2] VQI: Video Quality Indicators for DTV, [online], http://www2.um.edu.uy/ingenieria/vqi/, accessed: 9 June 2014.

[3] V. Q. E. G. (VQEG). [Online]. Available: http://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx

[4] Recommendation ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," Jan. 2012.

[5] Recommnedation ITU-R BT.710-4, "Subjective assessment methods for image quality in high definition television," Nov. 1998.

[6] Recommendation ITU-T P.910, "Subjective video quality assessment methods," Apr. 2008.

[7] Recommendation ITU-T P.913, "Methods for the subjective assessment of video quality, audio quality and audiovisual quality of internet video and distribution quality television in any environment," Jan. 2014.

[8] M. Pinson, "The consumer digital video library [best of the web]," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 172–174, 2013.

[9] K. Fliegel and C. Timmerer, "Qualinet multimedia database enabling qoe evaluations and benchmarking," *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)*, Mar. 2013. [Online]. Available: http://dbq-wiki.multimediatech.cz/_media/qi0306.pdf

[10] M. Leszczuk, L. Janowski, and M. Barkowsky, "Freely available large-scale video quality assessment database in full-hd resolution with h.264 coding," in *IEEE Globecom Workshops (GC Wkshps)*, Dec. 2013, pp. 1162–1167.

[11] M. Barkowsky, E. Masala, G. Van Wallendael, K. Brunnström, N. Staelens, and P. Le Callet, "Objective video quality assessment—towards large scale video database enhanced model development," *IEICE Transactions on Communications*, vol. 98, no. 1, pp. 2–11, 2015.

[12] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 165–182, June 2011.

[13] Y. Chen, K. Wu, and Q. Zhang, "From qos to qoe: A tutorial on video quality assessment," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 2, pp. 1126–1165, 2015.

[14] H. Wu and K. R. Rao, *Digital video image quality and perceptual coding*. CRC press, 2005.

[15] A. Takahashi, D. Hands, and V. Barriac, "Standardization activities in the itu for a qoe assessment of iptv," *IEEE Communications Magazine*, vol. 46, no. 2, pp. 78–84, Feb. 2008.

[16] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of jpeg compressed images," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 1, 2002, pp. 477–480.

[17] K. Yamagishi and T. Hayashi, "Qrp08-1: Opinion model for estimating video quality of videophone services," in *IEEE Global Telecommunications Conference (Globecom)*, Nov 2006, pp. 1–5.

[18] Recommendation ITU-T G.1070, "Opinion model for video-telephony applications," Apr. 2007.

[19] J. Joskowicz, R. Sotelo, and J. C. L. Ardao, "Towards a general parametric model for perceptual video quality estimation," *IEEE Transactions on Broadcasting*, vol. 59, no. 4, pp. 569–579, Dec. 2013.

[20] S. Winkler and P. Mohandas, "The evolution of video quality measurement: From psnr to hybrid metrics," *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 660–668, Sept. 2008.
[21] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, "New full-reference quality metrics based on hvs," in *CD-ROM proceedings of the second international workshop on video processing and quality metrics*, vol. 4, Scottsdale, USA, 2006.
[22] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of dct basis functions," in *Proceedings of the third international workshop on video processing and quality metrics*, vol. 4, 2007.
[23] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[24] Z. Wang and X. Shang, "Spatial pooling strategies for perceptual image quality assessment," in *IEEE International Conference on Image Processing*, Oct. 2006, pp. 2945–2948.
[25] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Conference Record of the Thirty-Seventh IEEE Asilomar Conference on Signals, Systems and Computers*, vol. 2, Nov. 2003, pp. 1398–1402.
[26] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, May. 2011.
[27] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
[28] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, "On advances in statistical modeling of natural images," *Journal of mathematical imaging and vision*, vol. 18, no. 1, pp. 17–33, 2003.
[29] M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky, "Random cascades on wavelet trees and their use in analyzing and modeling natural images," *Applied and Computational Harmonic Analysis*, vol. 11, no. 1, pp. 89–123, 2001.
[30] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp. 312–322, Sept. 2004.
[31] Recommendation ITU-T J.144, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," 2004.
[32] Recommendation ITU-R BT.1683, "Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference," 2004.
[33] S. Wolf and M. H. Pinson, "Low bandwidth reduced reference video quality monitoring system," in *First International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2005, pp. 23–25.
[34] VQEG, "Validation of Reduced-Reference and No-Reference Objective Models for Standard Definition Television, Phase I," 2009.
[35] Recommendation ITU-T J.249, "Perceptual video quality measurement techniques for digital cable television in the presence of a reduced reference," 2010.
[36] A. M. Rohaly, J. Libert, P. Corriveau, A. Webster *et al.*, "Final report from the video quality experts group on the validation of objective models of video quality assessment," *ITU-T Standards Contribution COM*, pp. 9–80, 2000.
[37] Recommendation ITU-T J.343, "Hybrid perceptual bitstream models for objective video quality measurements," 2014.
[38] M. Leszczuk, M. Hanusiak, I. Blanco, A. Dziech, J. Derkacz, E. Wyckens, and S. Borer, "Key indicators for monitoring of audiovisual quality," in *IEEE 22nd Signal Processing and Communications Applications Conference (SIU)*, 2014, pp. 2301–2305.
[39] M. Leszczuk, M. Hanusiak, M. Farias, E. Wyckens, and G. Heston, "Recent developments in visual quality monitoring by key performance indicators," *Multimedia Tools and Applications*, pp. 1–23, 2014.
[40] Recommendation ITU-R BT.2026, "Guidelines on the implementation of systems for in-service objective measurement and monitoring of perceptual transparency for the distribution chain of sdtv and hdtv programmes," Aug. 2012.
[41] Recommendation ITU-R BT.1885, "Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a reduced bandwidth reference," 2012.
[42] Recommendation ITU-R BT.1908, "Objective video quality measurement techniques for broadcasting applications using hdtv in the presence of a reduced reference signal," 2012.
[43] R. Sotelo, J. Joskowicz, J. P. Garella, D. Durán, and M. Juayek, "Subjective video quality test: Methodology, database and experience," in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, June 2015, pp. 1–6.
[44] J. Joskowicz, R. Sotelo, M. Juayek, D. Durán, and J. P. Garella, "Automation of subjective video quality measurements," in *Proceedings of the Latin America Networking Conference on LANC 2014*. ACM, 2014, p. 7.
[45] S. Péchard, R. Pépion, and P. Le Callet, "Suitable methodology in subjective video quality assessment: a resolution dependent paradigm," in *International Workshop on Image Media Quality and its Applications, IMQA2008*, Kyoto, Japan, Sep. 2008, p. 6. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00300182
[46] FFmpeg, [online], http://www.ffmpeg.org.
[47] J. Joskowicz and R. Sotelo, "A model for video quality assessment considering packet loss for broadcast digital television coded in h.264," *International Journal of Digital Multimedia Broadcasting*, vol. 2014, p. 1–11, 2014. [Online]. Available: http://dx.doi.org/10.1155/2014/242531
[48] J. Joskowicz, "Desarrollo de un modelo paramétrico general de estimación de la calidad percibida de video," Ph.D. dissertation, Ingeniería Telemática, Universidad de Vigo, Oct. 2012.
[49] E. Wu, Y. Diao, and S. Rizvi, "High-performance complex event processing over streams," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 2006, pp. 407–418.
[50] S. Winkler, "Analysis of public image and video databases for quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 616–625, Oct. 2012.
[51] J. Joskowicz, R. Sotelo, J. P. Garella, P. Zinemanas, and M. Simón, "Combining full reference and no reference models for broadcast digital tv quality monitoring in real time," *IEEE Transactions on Broadcasting*, vol. 62, no. 4, pp. 770–784, Dec 2016.
[52] J. P. Garella, E. Grampín, R. Sotelo, J. Baliosian, J. Joskowicz, G. Guimerans, and M. Simon, "Monitoring qoe on digital terrestrial tv: A comprehensive approach," in *2016 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, June 2016, pp. 1–6.
[53] A. T. Nasrabadi, M. A. Shirsavar, A. Ebrahimi, and M. Ghanbari, "Investigating the psnr calculation methods for video sequences with source and channel distortions," in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, June 2014, pp. 1–4.
[54] VQEG Final Draft Report, Validation of objective models of video quality assessment, Phase II, Mar. 2003 .

Juan Pablo Garella (M'15) received the Electrical Engineering degree from the Universidad de la República (UdelaR), Uruguay, in 2011 and it's Masters degree in Electrical Engineering at rhe same University, where he has been a Research Assistant since 2104. He participated in research projects on digital television with emphasis in perceived video quality estimation, full reference and no reference metrics, and QoS monitoring. His research interests include image processing, perceived video quality, quality of experience, digital TV, and ISDB-Tb.