



UNIVERSIDAD DE LA REPÚBLICA
FACULTAD DE INGENIERÍA



Modelado de calidad percibida de video en Televisión Digital Abierta

TESIS PRESENTADA A LA FACULTAD DE INGENIERÍA DE LA
UNIVERSIDAD DE LA REPÚBLICA POR

Juan Pablo Garella Lena

EN CUMPLIMIENTO PARCIAL DE LOS REQUERIMIENTOS
PARA LA OBTENCIÓN DEL TÍTULO DE
MAGISTER EN INGENIERÍA ELÉCTRICA.

DIRECTORES DE TESIS

Dr. José Joskowicz Zilberfarb Universidad de la República
Dr. Rafael Guzmán Sotelo Bovino..... Universidad de la República

TRIBUNAL

Dr. Federico La Rocca Universidad de la República
Dr. Alvaro Pardo..... Universidad Católica del Uruguay
Dr. Pedro Casas AIT Austrian Institute of Technology
Prof. Ing. María Simon..... Universidad de la República

DIRECTOR ACADÉMICO

Dr. Rafael Guzmán Sotelo Bovino..... Universidad de la República

Montevideo
sábado 1 abril, 2017

Modelado de calidad percibida de video en Televisión Digital Abierta, Juan Pablo
Garella Lena.

ISSN 1688-2806

Esta tesis fue preparada en L^AT_EX usando la clase iietesis (v1.1).

Contiene un total de 135 páginas.

Compilada el sábado 1 abril, 2017.

<http://iie.fing.edu.uy>

“Es justo allí a mitad de camino entre el huerto desnudo y el huerto verde, cuando las ramas están a punto de estallar en flor, en rosa y blanco, que tememos lo peor.”

ROBERT FROST. ESCRITOR (1874-1963)

Agradecimientos

A mis tutores Rafael Sotelo y José Joskowicz, por abrirme las puertas al mundo de la investigación y ofrecerme la oportunidad de trabajar en el proyecto titulado: “Indicadores de Calidad de Video”(VQI), ANII-FST 1_2012_1_8147 patrocinado por MIEM-DINATEL y gestionado por la Fundación Julio Ricaldoni (FJR).

A María Simon, por la oportunidad de participar en el desarrollo del Protocolo de Homologación de Receptores ISDB-Tb y en el proyecto titulado: “Sistema de Monitorización de la señal de TV Digital”(SMTVD), ANII - FST_1_2013_1_13436 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).

A la Dirección Nacional de Telecomunicaciones y Servicios de Comunicación Audiovisual del Ministerio de Industria Energía y Minería (MIEM-Dinatel) por los fondos destinados a los convenios y proyectos de investigación en Televisión Digital Abierta.

A la Agencia Nacional de Investigación e Innovación (ANII) por concederme la oportunidad de obtener la beca de Posgrados Nacionales.

A Anderson Fagiani, por compartir su experiencia respecto de la norma ISDB-T.

A mi familia y amigos, Pablo Flores Guridi amigo y colega, Gustavo Guimerans, soldado, Martín Cáceres (Paso), Juan Pablo Braga y por último, pero no menos importante, al Gallinero y a su gente.

A María Cecilia.

Resumen

En esta tesis se llega al estado del arte en evaluación de calidad de video en su modalidad subjetiva y objetiva. Se evalúan y analizan distintos modelos objetivos de estimación de calidad, con referencia completa, referencia reducida y sin referencia, propuestos previamente por el grupo de investigación, por la comunidad académica o estandarizados por organismos internacionales. Se compara su desempeño contrastando los resultados obtenidos por medio de la realización de evaluaciones subjetivas de calidad de video. Se profundiza en el estándar ISDB-Tb de Televisión Digital Abierta, desde el punto de vista de la recepción y visualización de la señal de Televisión Digital. Por último se propone y evalúa un novedoso método que combina modelos objetivos con referencia completa y sin referencia para efectuar la medición de calidad percibida de video en tiempo real con aplicación en Televisión Digital Abierta.

Tabla de contenidos

Agradecimientos	III
Resumen	VII
1. Introducción	1
1.1. Motivación	2
1.2. Participación en Proyectos	3
1.3. Objetivos de la tesis	5
1.4. Contribuciones Principales	6
1.5. Organización del documento	6
2. Estado del Arte en Evaluación de Calidad de Video	7
2.1. Introducción	7
2.2. Evaluación Subjetiva de Calidad	7
2.3. Evaluación Objetiva de Calidad	12
2.4. Video Quality Experts Group VQEG	23
2.5. Técnicas de desempeño y comparación de modelos	27
3. Televisión Digital Terrestre: norma ISDB-Tb	29
3.1. Introducción	29
3.2. Codificación de fuente	30
3.2.1. Codificación de video	31
3.2.2. H.264/AVC	36
3.2.3. Generación del flujo de transporte MPEG-2 TS	37
3.3. Codificación de canal	42
4. Evaluación Subjetiva de la calidad percibida de video en TV Digital	47
4.1. Introducción	47
4.2. Descripción general de la base de datos	47
4.3. Selección y preparación del material	48
4.4. Metodología de evaluación	54
4.5. Sistema de automatización de pruebas subjetivas	56
4.6. Análisis de resultados subjetivos	57

Tabla de contenidos

5. Aplicación de un modelo paramétrico de estimación de calidad de video en TV Digital	65
5.1. Introducción	65
5.1.1. I_c - Degradación asociada al proceso de codificación	66
5.1.2. I_p - Degradación de calidad asociada a errores en transmisión	69
5.1.3. MOSp - Predicción del MOS	73
6. Evaluación de calidad de video en un sistema de monitorización de la señal de TV Digital	77
6.1. Introducción	77
6.2. Flujo de trabajo de una estación de Televisión: <i>puntos de control de calidad</i>	78
6.3. Sistema de Monitorización de la señal de TV Digital	80
6.4. Modelo global de estimación de calidad de video	82
6.4.1. Cálculo de I_c	83
6.4.2. Cálculo de I_p	84
6.5. MOSp - Predicción del MOS	91
6.6. Alineación temporal de clips de video	93
6.7. Indicadores MOAVI	97
7. Conclusiones	103
Apéndices	107
A. Publicaciones realizadas en el marco de este trabajo	107
Referencias	109
Índice de tablas	117
Índice de figuras	118

Capítulo 1

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 como VQA (*Video Quality Assessment*). Dependiendo del tipo de señal se pueden identificar otras ramas, como IQA (*Image Quality Assessment*) o evaluación de calidad de imágenes, extendiéndose a todo tipo de estímulo 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

Capítulo 1. Introducción

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. Se presenta la norma ISDB-T de TVD adoptada por Uruguay y se describe la realización de evaluaciones subjetivas con el fin de confeccionar una base de datos para 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 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 de la presente tesis, su marco de trabajo, aportes específicos en el área de evaluación de calidad percibida de video, con su respectiva aplicación en TVD.

1.1. 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

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

1.2. Participación en Proyectos

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.

1.2. Participación en Proyectos

Acompañando la implantación de la TVD en el territorio uruguayo el gobierno ha incentivado a través de la Dirección Nacional de Telecomunicaciones y Servicios

Capítulo 1. Introducción

de Comunicación Audiovisual (DINATEL) ² y de la Agencia Nacional de Investigación e Innovación (ANII) ³ la realización de proyectos vinculados a la temática. En el transcurso de esta tesis de maestría he participado activamente en varios de estos proyectos. A continuación se describen cada uno de estos y dado que su ejecución fue de carácter colectivo se especifica el rol asumido y la actividad principal desempeñada. 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 rol asumido en este proyecto fue el de investigador. Las tareas principales asociadas al cargo consistieron del estudio del estado del arte en modelos objetivos de evaluación de calidad de video sin referencia (NR), herramientas de análisis del flujo de información recibido de la señal ISDB-T, herramientas de extracción de características del video decodificado, activa participación en la realización de evaluaciones subjetivas de calidad de video, análisis de las degradaciones asociadas al proceso de compresión de video y transmisión de la señal ISDB-T.

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. Al igual que en el proyecto anterior, en este proyecto se tuvo un rol de investigador. La principal actividad desempeñada tuvo lugar en el área de calidad de la señal, particularmente en la calidad percibida de video. Se asumieron las siguientes tareas, estudio de modelos FR, RR y NR de evaluación de calidad de video con el fin de estimar las degradaciones asociadas al proceso de adquisición, codificación y errores en transmisión

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

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

1.3. Objetivos de la tesis

y luego su posterior implementación e integración al sistema de monitorización de la señal de TVD.

Por otro lado, he participado de la confección del *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 del Uruguay (LATU) para ensayar receptores acorde a lo especificado en el Decreto 143/013 del Ministerio de Industria Energía y Minería, Poder Ejecutivo. En este proyecto se tuvo participación directa en todo lo asociado al diseño, desarrollo de los conjuntos de prueba, transferencia de conocimiento y puesta a puntos de los ensayos directamente en las facilidades donde actualmente el protocolo es aplicado.

Por último he participado en la escuela de verano titulada: “*Statistical and learning methods for reliable assessment of Quality of Experience*”, llevada a cabo en el programa: Qualinet Summer School 2014 ⁵ en Polytech Nantes, École polytechnique de l’université de Nantes, Francia. El trabajo realizado, en el cual tuve participación directa, por el grupo de investigadores formado en esta escuela se materializó en la publicación presentada en [3] relacionado a calidad de video y con aplicación directa sobre esta tesis. La actividad principal desarrollada fue en el área de evaluación subjetiva de calidad, extracción de características de secuencias de video y cómputo de modelos objetivos FR sobre conjuntos de secuencias.

Es en estos proyectos que se enmarca el presente trabajo.

1.3. Objetivos de la tesis

En lo que sigue se especifican los objetivos específicos de esta tesis:

- Estudio del estado del arte con foco en modelos objetivos para evaluación de calidad de video con referencia completa (FR), de referencia reducida (RR) y sin referencia (NR).
- Procesamiento de video: estudio de herramientas de extracción de características del contenido y flujo de bits de información (bitstream) sobre clips de video registrados de TVD.
- Análisis de resultados y activa participación en la realización de evaluaciones subjetivas de calidad de video.
- Análisis de las degradaciones asociadas al proceso de compresión de video (compresión H.264/AVC) y al proceso de transmisión de la señal de TVD.
- Análisis y selección de un modelo FR o RR para el cálculo de las degradaciones asociadas al proceso de transmisión.

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

⁵Enlace: <https://sites.google.com/site/3rdqualinetsummerschool2014/home>

Capítulo 1. Introducción

- Análisis y selección de un modelo NR para el cálculo de las degradaciones asociadas al proceso de adquisición y codificación.
- Implementación e integración de los modelos seleccionados para su cómputo en tiempo real a un sistema de monitorización de la señal de TVD.

1.4. Contribuciones Principales

Esta tesis contiene aportes específicos en el área de evaluación objetiva de calidad de video con aplicación en Televisión Digital Abierta. Se propone y evalúa un nuevo método que combina modelos objetivos con referencia completa (FR) y sin referencia (NR) para efectuar la medición de calidad percibida de video en tiempo real para la señal de TVD. Esta nueva aproximación se basa en la utilización de técnicas modernas de evaluación de calidad de video bajo un sistema de monitorización con nodos receptores de TVD distribuidos geográficamente en el área de cobertura de una estación de TVD.

1.5. Organización del documento

En el capítulo 2 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 el capítulo 3 se presenta el estándar ISDB-T de Televisión Digital Abierta adoptado por Uruguay. En el capítulo 4 se describe la metodología, resultados y experiencia obtenida en la realización de evaluaciones subjetivas de calidad de video. En el capítulo 5 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 el capítulo 6 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 el capítulo 7 se presentan las conclusiones del trabajo y líneas de investigación a futuro. Finalmente, en el Apéndice A se mencionan las publicaciones realizadas en el marco de esta tesis.

Capítulo 2

Estado del Arte en Evaluación de Calidad de Video

2.1. Introducción

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) [4] especialmente en materia de televisión y se describen los indicadores generalmente utilizados para comparar el desempeño entre modelos objetivos de estimación de calidad.

2.2. Evaluación Subjetiva de Calidad

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 [5] fue especialmente desarrollada para aplicaciones de Televisión en definición estándar (SD), mientras que la recomendación ITU-R BT.710-4 [6] extiende su

Capítulo 2. Estado del Arte en Evaluación de Calidad de Video

uso para alta definición (HD). Por otro lado, la recomendación ITU-T P.910 [7] describe metodologías de evaluación subjetiva para aplicaciones multimedia. Recientemente, en 2014, fue aprobada la recomendación ITU-T P.913 [8]; 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.

Generalmente con el fin de realizar pruebas subjetivas estas metodologías requieren de un proceso común, este proceso se puede resumir como sigue:

1. Selección apropiada del material (clips de video)
2. Preparación de un conjunto de clips de video controlado
3. Preparación de un ambiente de pruebas (pantalla, iluminación, distancia de visión, etc.)
4. Definición del método de evaluación y preparación de los formularios o dispositivos de adquisición para las calificaciones (típicamente sobre papel o con un dispositivo electrónico)
5. Reclutamiento de espectadores no expertos
6. Proceso de despliegue de los clips de video a los espectadores en orden aleatorio, normalmente en muchas sesiones diferentes
7. Introducción de las evaluaciones resultantes en una hoja de cálculo o software específicamente diseñado para tal fin
8. Realización de un análisis estadístico sobre los resultados (*post screening*)

Dentro de los métodos de evaluación, a continuación se describen los más utilizados para evaluar calidad subjetiva de video.

Método de escala de degradación con doble estímulo (DSIS, *Double Stimulus Impairment Scale*)

En el método DSIS [5], 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 una escala de cinco puntos, esta escala se puede ver en la tabla 2.1. A este método también se lo conoce como Índices por Categorías de Degradación (DCR, *Degradation Category Rating*) [5].

2.2. Evaluación Subjetiva de Calidad

Tabla 2.1: Escala discreta de cinco niveles, (DSIS)

5	Imperceptible
4	Perceptible, pero no molesta
3	Ligeramente molesta
2	Molesta
1	Muy molesta

Método de escala de calidad continua de doble estímulo (DSCQS, *Double Stimulus Continuous Quality Scale*)

El método DSCQS [5] es similar al método DSIS. Al espectador se le presentan dos clips de video, el de referencia y el degradado pero no se le dice cuál es el original, puede ser el primero o el segundo. A los espectadores simplemente se les pide evaluar la calidad de cada clip de video mediante la inserción de una marca en un escala vertical continua. Las escalas verticales se imprimen en pares como se muestra en la figura 2.1.

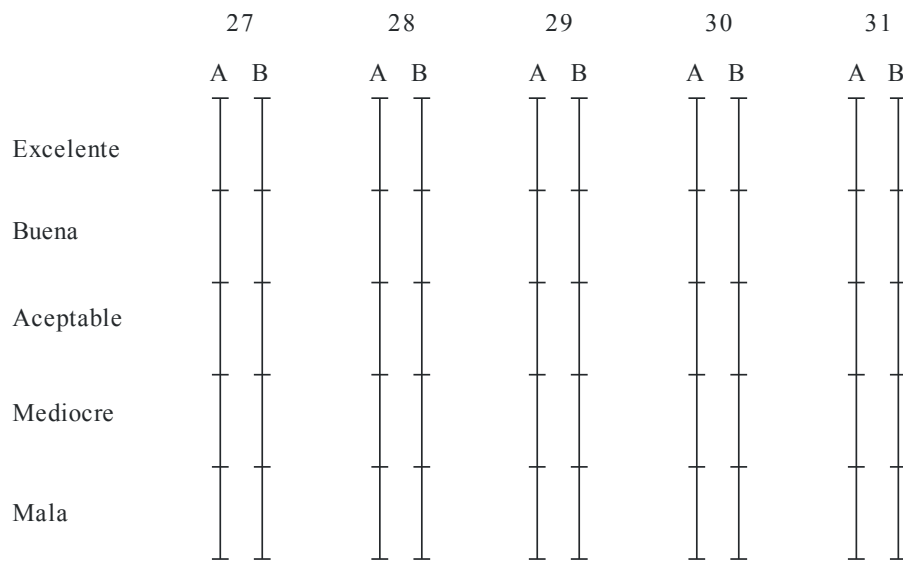


Figura 2.1: Ejemplo de escalas verticales impresas para evaluación subjetiva de la calidad mediante el método DSCQS.

Evaluación de calidad continua de estímulo único (SSCQE, *Single Stimulus Continuous Quality Evaluation*)

En el método SSCQE [5] la calidad de video se mide de manera continua, donde los sujetos observan el material solo una vez sin una referencia. Pueden probarse secuencias más largas, asemejándose a un servicio real. Las recomendaciones citadas sugieren el uso de un dispositivo electrónico conectado a un ordenador para el

Capítulo 2. Estado del Arte en Evaluación de Calidad de Video

registro de la evaluación continua de la calidad por parte de los espectadores.

Método de doble estímulo simultáneo para evaluación continua (SDSCE, *Simultaneous Double Stimulus for Continuous Evaluation*)

El principal cambio que ofrece el método SDSCE [5] es la presentación del video de referencia simultáneamente al video bajo ensayo. Dependiendo de la resolución de los videos, las secuencias se pueden visualizar en la misma pantalla, si no es el caso se deben utilizar dos pantallas alineadas. Los sujetos tienen que votar en una escala de 0 a 100, siendo 100 perfecto en función de la similitud y 0 cuando la fidelidad es nula. Además los sujetos conocen cual de los videos es el de referencia y se les pide evaluar su calidad de manera continua mientras se despliegan las secuencias.

Índices por categorías absolutas (ACR, *Absolute Category Rating*)

En el método ACR [7] 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 2.2.

Tabla 2.2: Escala discreta de cinco niveles, (ACR)

5	Excelente
4	Buena
3	Aceptable
2	Mediocre
1	Mala

Una variante a este método se denomina por sus siglas en ingles como ACR-HR [5] (*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.

Método de comparación por pares (PC, *Pair Comparison*)

En el método PC [7] las secuencias de prueba se presentan en parejas, se utiliza la misma secuencia con dos degradaciones diferentes y los espectadores deben elegir su preferida en el par. Los pares deben estar presentados en todas las combinaciones posibles. Al igual que en otros métodos, si la resolución utilizada es pequeña, cada par de secuencias se puede mostrar simultáneamente en el mismo monitor.

Los métodos *DSIS*, *DSCQS*, *ACR* y *PC* históricamente fueron utilizados con formularios de papel además de utilizar un dispositivo de reproducción sin posibilidad de pausa, imponiendo una restricción a los espectadores de evaluar el

2.2. Evaluación Subjetiva de Calidad

estímulo presentado con un retardo fijo (ver, por ejemplo [9]). Para los métodos *SSCQE* y *SDSCE* las recomendaciones sugieren el uso de dispositivos electrónicos especiales con el fin de realizar una evaluación continua a lo largo de la duración del clip de video. Alternativamente, para todos los casos, se utiliza a menudo una aplicación basada en software, proporcionando una escala para hacer la evaluación con el ratón o teclado. En [10] se realiza una comparación entre cuatro escalas distintas, en este caso los espectadores brindan sus evaluaciones usando un ratón de computadora en distintas escalas, continuas y discretas. Del mismo modo en [11] luego de terminada la presentación del video de prueba, se presenta en pantalla una escala de cinco puntos y se le pide al espectador que califique el video con el ratón o el teclado. En [12] se realiza una discusión sobre la idoneidad de diversos dispositivos utilizados para capturar las calificaciones del espectador en donde se presentan cuatro tipos de interfaz: ratón, joystick, acelerador y barra deslizante. Otro tipo de enfoque se presenta en [13] donde las calificaciones brindadas por los usuarios al evaluar la calidad de clips de video en dispositivos móviles se adquieren mediante el ángulo de giro de un volante de video juegos. Todos estos métodos tienen ciertas desventajas, los formularios en base a papel son tediosos para rellenar por parte de los espectadores, además pueden ocurrir errores al rellenarlos o al pasar el total de resultados a un sistema informático (hojas de cálculo o programas de análisis) aumentando la posibilidad de obtener conclusiones erróneas. Por otro lado los dispositivos electrónicos especiales que se proponen son difíciles de conseguir y su uso es específico para cada tipo de metodología. Además, solo uno o dos de estos dispositivos se pueden conectar simultáneamente a un ordenador, lo que limita el número de espectadores por sesión. Por otro lado la utilización de un ratón o teclado es especialmente tediosa para realizar evaluación continua de calidad [12]. Con el objetivo de superar estas desventajas y facilitar la realización de evaluaciones subjetivas de calidad de video, en este trabajo se participó en el diseño de un sistema basado en software que automatiza las pruebas y la recolección de datos. El funcionamiento del sistema y los ensayos subjetivos realizados se describen con mayor detalle en el capítulo 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) [14] ¹. 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 [15] se pueden encontrar varias referencias a bases de datos públicas con videos calificados. Además existen esfuerzos para construir y utilizar bases de da-

¹CDVL , [En línea], <http://www.cdvl.org>

tos de videos en gran escala [16] [17].

2.3. 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 [18]. 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 [19], 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: con referencia completa (FR), de referencia reducida (RR) y sin referencia (NR) [20]. Esta clasificación se basa en el grado de utilización de una referencia para el cómputo de la estimación de calidad. Los modelos con referencia completa utilizan el video original completo en conjunto con el degradado. Los modelos con referencia reducida son aquellos en donde es necesaria información parcial de la señal original y por último los modelos sin referencia son aquellos que no necesitan del video original. 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 [19]. 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 [21] se puede encontrar una tercer clasificación donde se encuentran los modelos paramétricos, que utilizan información del flujo de bits de información (*bitstream*), de los cabecales o la carga útil (*payload*) de los paquetes para computar el MOS. Además se describen modelos híbridos que combinan distintas

2.3. Evaluación Objetiva de Calidad

aproximaciones. Por último, adicionalmente a las clasificaciones anteriores, se denominan como modelos basados en píxeles o *pixel-based* a aquellos modelos con referencia completa que utilizan el contenido, valor de los píxeles, 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 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 [22] 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 [23] 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 [24]. Por otro lado en [25], se presenta un modelo paramétrico de estimación de calidad de video en aplicaciones de SDTV y HDTV. En este caso el modelo propuesto utiliza como parámetros el *bitrate* y el porcentaje de pérdida de paquetes no teniéndose en cuenta el contenido ni la tasa de cuadros. Posteriormente, en [26] se propone una extensión al modelo previamente comentado, donde se incluyen parámetros relacionados al contenido, aplicado a IPTV en resolución HD ateniéndose solo a degradaciones obtenidas en el proceso de codificación. En [27] y [28] se propone un modelo que considera como parámetros el *bitrate*, el *frame rate* y la pérdida de paquetes. Además en el trabajo se resalta la dependencia de la calidad en función del contenido. Siguiendo la evolución de los modelos paramétricos en [29] José Joskowicz et al. proponen un modelo paramétrico que combina los efectos del *bitrate*, *frame rate*, resolución de la pantalla y contenido del video. El modelo fue desarrollado para resoluciones SD, VGA, CIF y QCIF, y utiliza una métrica que caracteriza el tipo de contenido denominada por sus siglas en inglés como SAD (Sum of Average Differences). Además en [30] 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. 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

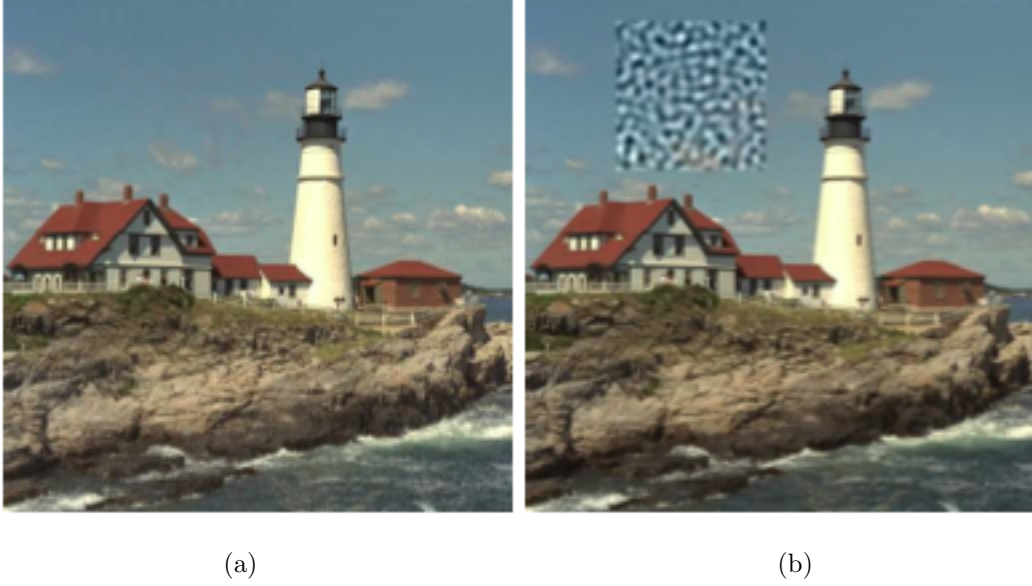


Figura 2.2: Visibilidad de distintos tipos de distorsión, ambas imágenes tienen el mismo PSNR aunque su calidad percibida sea distinta. Imagen extraída de [31].

secuencia de imágenes, permitiendo su aplicación para video.

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 2.1 y 2.2 respectivamente, en donde, a modo de ejemplo, se utiliza una imagen de tamaño $M \times N$ en escala de gris.

$$\text{MSE}(f, g) = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (f_{ij} - g_{ij})^2 \quad (2.1)$$

$$\text{PSNR}(f, g) = 10 \log_{10} \left(\frac{L_{max}^2}{\text{MSE}(f, g)} \right) \quad (2.2)$$

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. En la figura 2.2 se ilustra un ejemplo. Observando las imágenes de la figura 2.2(a) y 2.2(b) si bien ambas tienen el mismo $PSNR$ su calidad percibida es muy distinta.

2.3. Evaluación Objetiva de Calidad

Existen algunas variantes al tradicional PSNR, es el caso de los modelos PSNR-HVS [32] y PSNR-HVS-M [33] 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 [34]. 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:

$$SSIM(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma \quad (2.3)$$

Donde l representa la luminancia, c contraste y s estructura, a continuación se presenta su formulación:

$$l(x, y) = \frac{(2\mu_x\mu_y + C_1)}{(\mu_x^2 + \mu_y^2 + C_1)} \quad (2.4)$$

$$c(x, y) = \frac{(2\sigma_x\sigma_y + C_2)}{(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2.5)$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{(\sigma_x\sigma_y + C_3)} \quad (2.6)$$

Operando y realizando la siguiente sustitución: $C_3 = C_2/2$ y $\alpha = \beta = \gamma = 1$, recomendada por sus autores en [34], se obtiene la siguiente expresión:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2.7)$$

En donde,

Capítulo 2. Estado del Arte en Evaluación de Calidad de Video

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.8)$$

$$\sigma_x = \left(\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2 \right)^{\frac{1}{2}} \quad (2.9)$$

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y) \quad (2.10)$$

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 2.7 en una ventana móvil de tamaño $N \times N$, por ejemplo 8×8 , pixel a pixel a través de la imagen. En general, para aplicar el índice SSIM se utiliza una ventana *Gaussiana* circular, simétrica y normalizada, de tamaño 11×11 [34]. 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:

$$MSSIM(X, Y) = \frac{1}{M} \sum_{j=1}^M SSIM(x_j, y_j) \quad (2.11)$$

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. A continuación, en la figura 2.3 se pueden observar distintas degradaciones de una misma imagen original manteniendo constante el MSE. El modelo MSSIM ofrece una mejor correlación con la calidad percibida, siendo 1 cuando las imágenes son iguales (no hay degradación) y 0 en caso contrario.

En la figura 2.4 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

2.3. Evaluación Objetiva de Calidad

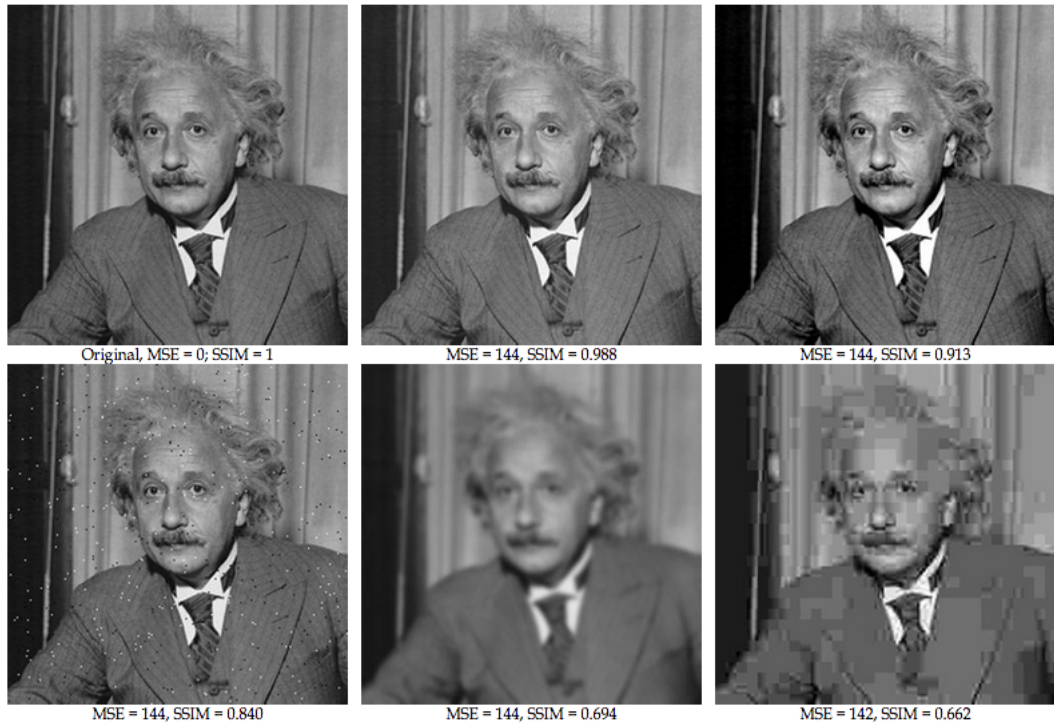


Figura 2.3: Imágenes degradadas con igual MSE. Se aplica la métrica MSSIM obteniéndose una mejor correlación con la calidad percibida.

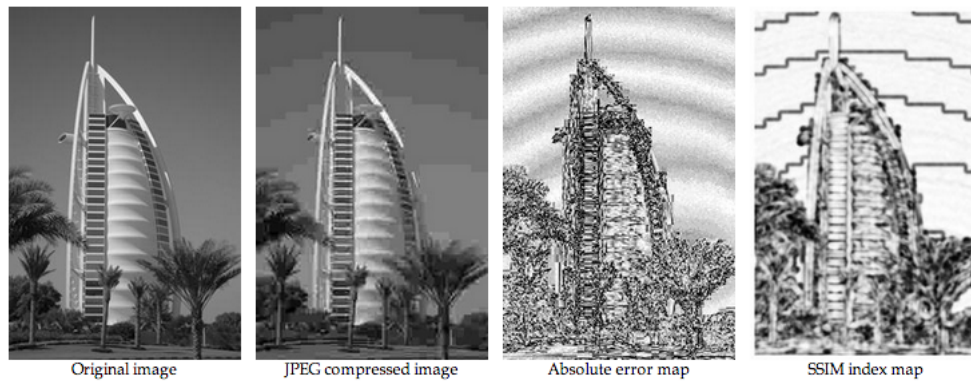


Figura 2.4: Imagen Original, Imagen degradada, Absolute Error Map, SSIM Index Map. Imagen extraída de [35].

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 [35].

Con el paso de los años la métrica MSSIM, ha evolucionado. Por ejemplo, la métrica MS_SSIM [36] 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. El procedimiento consiste básica-

Capítulo 2. Estado del Arte en Evaluación de Calidad de Video

mente en tomar la imagen de referencia y la degradada, luego a cada una se le aplica un filtro pasa bajos y posteriormente se submuestran los pixeles de la imagen reduciendo la altura y el ancho resultante a la mitad. Este proceso se repite M veces, en cada caso al resultado se le aplica el índice SSIM (ecuaciones: 2.4, 2.5, 2.6), los que luego son utilizados para generar un índice de calidad global de la imagen, como se muestra en la siguiente expresión.

$$MS_SSIM(X, Y) = [l_M(x, y)]^{\alpha_M} \prod_{j=1}^M [c_j(x, y)]^{\beta_j} [s_j(x, y)]^{\gamma_j} \quad (2.12)$$

En donde l corresponde a la luminancia, c al contraste y s a estructura, los subíndices j identifican a la escala siendo $j = 1$ la de la imagen original, α_M , β_j , γ_j son parámetros de calibración.

Continuando con la evolución del índice SSIM existen distintas variantes. Por ejemplo, en [35] 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 [37]. 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 [37]. De esta manera además generan una variante al PSNR denominada IW-PSNR.

Criterio de Fidelidad Visual de la Información (VIF, *Visual Information Fidelity*)

El modelo VIF [38] 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 [39] 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 [40].

En la figura 2.5 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

2.3. Evaluación Objetiva de Calidad

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 2.13

$$VIF = \frac{\sum_{j=0}^S \sum_{i=1}^{M_j} I(c_{i,j}; f_{i,j})}{\sum_{j=0}^S \sum_{i=1}^{M_j} I(c_{i,j}; e_{i,j})} \quad (2.13)$$

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 específica 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.

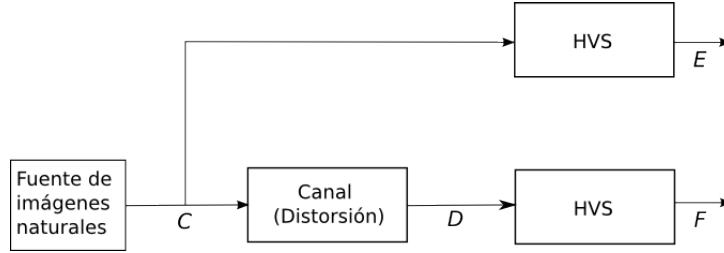


Figura 2.5: La información mutua entre la señal C y E cuantifica la información que el cerebro idealmente puede extraer de la imagen de referencia. Por otro lado la información mutua entre la señal C y F cuantifica lo que idealmente el cerebro puede extraer cognitivamente de la señal bajo ensayo.

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 una mejora lineal del contraste tiende a mejorar la calidad percibida en una imagen. Esta propiedad se puede ver en la figura 2.6, extraída de [38], donde a partir de una imagen de referencia 2.6 (a) se generan tres imágenes

Capítulo 2. Estado del Arte en Evaluación de Calidad de Video



(a) Imagen de referencia, VIF=1.0



(b) Imagen con mejora de contraste, VIF=1.1



(c) Imagen borrosa, VIF=0.07



(d) Imagen con compresión JPEG, VIF=0.1

Figura 2.6: Propiedades del modelo VIF: Imagen de referencia (a), con mejora de contraste (b), con compresión JPEG (c) y con borrosidad por medio de un filtro pasa bajos (d).

distorsionadas: con mejora de contraste 2.6 (b), con compresión JPEG 2.6 (c) y aplicando un filtro pasabajos con el objetivo de generar una imagen borrosa 2.6 (d). Se puede ver que el modelo VIF es sensible a los distintos tipos de distorsión y su impacto en la percepción de la calidad.

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 2.13, 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

2.3. Evaluación Objetiva de Calidad

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 [41] 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 [42] e ITU-R BT.1683 [43]. Este modelo, acorde con lo estipulado en [42] 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. Por ejemplo, si uno desea contrastar los resultados del modelo en una escala similar a la utilizada en evaluaciones subjetivas realizadas conforme a la metodología DSCQS basta multiplicar su resultado por 100. El modelo utiliza siete parámetros, cuatro de ellos se basan en características extraídas de los gradientes espaciales de la componente de luminancia Y , dos se basan en características extraídas del vector formado por las dos componentes de crominancia (C_B , C_R) y el restante se basa en las características de contraste y la información temporal absoluta de la componente de luminancia Y . La selección de éstos se determinó por el criterio de maximizar la correlación entre los resultados de calidad percibida por el modelo y los resultados subjetivos de calidad.

En [42] se define una característica de calidad como la cantidad de información asociada a una subregión espacio-temporal (S-T, *Spatial-Temporal*) de una secuencia de video (original o procesada) o extraída de ésta. En tanto las secuencias de características son función del espacio y del tiempo. Comparando las características extraídas del video procesado (degradado) calibrado con las extraídas del video original es que se calculan los parámetros de calidad indicativos de las variaciones perceptuales de la calidad de video. Con el objetivo de extraer estas características para ser utilizadas por el modelo, en primer lugar se aplica un filtro perceptual a la secuencia de video. Esto se realiza con el objetivo de realzar algunas propiedades, por ejemplo la información de borde. Luego se subdivide la secuencia en subregiones espacio-temporales (S-T) y se extraen las características de estas subregiones. Por último, se aplica un umbral de perceptibilidad a las características extraídas. Una región S-T, por ejemplo, puede ser de 8 píxeles horizontales x 8 líneas verticales x 6 cuadros de video NTSC, que corresponde a un total de 384

Capítulo 2. Estado del Arte en Evaluación de Calidad de Video

pixeles. Cuando se aplica a video de 25 fps (PAL), esta misma región S-T abarca 8 pixeles horizontales x 8 líneas verticales x 5 cuadros de video, lo que corresponde a un total de 320 pixeles.

A continuación se describen brevemente las principales características que utiliza el modelo VQM *General*.

Características basadas en gradientes espaciales

Con el fin de caracterizar las distorsiones perceptuales en los bordes se utilizan características derivadas de gradientes espaciales. Las componentes Y de las secuencias de video original y procesada se filtran por medio de filtros de realce de bordes horizontales y verticales. Luego estas secuencias de video filtradas se dividen en regiones espacio-temporales (S-T) de las que se extraen las características que cuantifican la actividad espacial en función de la orientación angular. Por último, se aplica un umbral de perceptibilidad.

Características basadas en la información de crominancia

Respecto de las características basadas en la información de crominancia (C_b , C_r), se calcula un vector que representa el valor medio de las componentes de color en una región S-T, en el caso de las componente C_r se pondera con una constante de 1,5, según con la siguiente ecuación:

$$f_{cohercolor} = (C_b, 1,5C_r) \quad (2.14)$$

Siendo C_b y C_r el valor medio de las componentes de crominancia en una región S-T. En general se utiliza una región de 8 pixeles horizontales x 8 líneas verticales x (1 a 3) cuadros de video, que en realidad disminuye su tamaño dado el submuestreo de la señal de crominancia, por ejemplo a 4 pixeles horizontales acorde a la recomendación ITU-R BT.601-5 [44].

Características basadas en la información de contraste

En cuanto las características utilizadas que miden información de contraste localizada, éstas son sensibles a degradaciones de calidad tales como borrosidad (pérdida de contraste) o ruido aditivo (ganancia de contraste). Se calcula para una región S-T tomando la desviación estándar de las componentes de la señal de luminancia Y dentro de la región y se denomina f_{cont} .

Características basadas en información temporal absoluta (ATI, *Absolute Temporal Information*)

Este tipo de características, que miden distorsiones en el flujo de movimiento, son sensibles a degradaciones de calidad tales como la omisión o repetición de cuadros (pérdida de movimiento) o ruido aditivo (ganancia de movimiento). La característica de información temporal absoluta denominada f_{ATI} se calcula para cada región S-T. En donde se genera una nueva secuencia de video que es el valor

2.4. Video Quality Experts Group VQEG

absoluto de la diferencia entre los cuadros de video consecutivos en los instantes t y $t - 1$. A continuación se calcula la desviación estándar en la región S-T y por último se aplica un umbral de perceptibilidad.

Características basadas en el producto vectorial del contraste y la información temporal absoluta

El nivel de afectación de calidad percibida debido a degradaciones espaciales puede verse influido por la cantidad de movimiento presente en la escena. Análogamente la percepción de degradaciones temporales puede verse influida por la cantidad de detalle. En tanto, con el objetivo de tomar en cuenta esta interacción se toma una característica que se deriva del producto vectorial entre la información de contraste y la información temporal absoluta. Al utilizar esta formulación las degradaciones en las regiones S-T serán más visibles cuando el producto vectorial sea bajo (por ejemplo, baja actividad espacial y temporal) y menos visibles cuando es alto (por ejemplo, alta actividad espacial y temporal). Esto es especialmente cierto para degradaciones como el ruido y los bloques con errores.

Finalmente los parámetros de calidad que surgen a partir de estas características se calculan en regiones S-T de las señales originales y procesadas, a continuación se agrupan las degradaciones utilizando una función de agrupamiento espacial y luego temporal. Como resultado se tiene un único parámetro de cada característica para toda la secuencia. Finalmente se realiza una combinación lineal para obtener una estimación de la calidad percibida.

VQM *LowBw* NTIA

Una de las variantes propuestas al modelo VQM *General* es el modelo VQM *Lowbw* [45] 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 [46]. 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 [47].

2.4. Video Quality Experts Group VQEG

El VQEG [4] lleva a cabo un gran trabajo sistemático y objetivo de comparación de modelos. El objetivo del VQEG es proporcionar evidencia a los organismos internacionales de estandarización acerca del desempeño o rendimiento de diversos modelos propuestos, a los efectos de definir modelos estándar y objetivos de calidad

Capítulo 2. Estado del Arte en Evaluación de Calidad de Video

percibida de video. El VQEG ha trabajado en la evaluación de modelos basados en el análisis de contenidos, del tipo FR, RR y NR, para diferentes aplicaciones.

Los proyectos finalizados por el VQEG han sido los siguientes:

- *FRTV Phase I* (Finalizado: Junio, 2000)
- *FRTV Phase II* (Finalizado: Agosto, 2003)
- *Multimedia Phase I* (Finalizado: Setiembre, 2008)
- *RRNR-TV* (Finalizado: Junio, 2009)
- *HDTV Phase I* (Finalizado: Junio, 2010)
- *Hybrid Perceptual/Bitstream* (Finalizado: Setiembre, 2014)

En cada proyecto se han evaluado modelos específicos para ciertas aplicaciones, según lo indican sus nombres. Algunos de los proyectos del VQEG han terminado en recomendaciones aprobadas por el Sector de Estandarización de la Unión Internacional de Telecomunicaciones (ITU-T). Por ejemplo, el proyecto *FRTV Phase I* fue el primer experimento de validación realizado por VQEG, donde se examinaron modelos objetivos con referencia completa (FR) y sin referencia (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 [48]. 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 con referencia completa (FR) y sin referencia (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. Como resultado se generan las siguientes recomendaciones: ITU-T Rec. J.144 (2004) [42] e ITU-R Rec. BT.1683 (2004) [43].

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. El modelo PSNR se utilizó como implementación de referencia de mínimo desempeño y fue estandarizada. Como resultado del proyecto, ITU realizó las siguientes estandarizaciones: ITU-T Rec. J.247 (2008) [49], ITU-T Rec. J.246 (2008) [50], ITU-R Rec. BT.1866 (2010) [51], ITU-R Rec. BT.1867 (2010) [52] y la ITU-T Rec J.340 (2010) [53].

2.4. Video Quality Experts Group VQEG

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) [47]. En el proyecto *HDTV Phase I* se examinaron modelos objetivos con referencia completa (FR), referencia reducida (RR) y sin referencia (NR) con el objetivo de predecir la calidad de la televisión de alta definición. Todos los modelos NR fueron descartados. Como resultado, ITU considero que el desempeño de algunos modelos FR y RR justificaba su estandarización por lo tanto se generaron las recomendaciones ITU-T Rec. J.342 [54] y ITU-T Rec. J.341 [55]. 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) [56]. En esta recomendación se destacan los siguientes modelos:

- *ITU-T Rec. J.343.1* (2014), “Hybrid-NRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of encrypted bitstream data”
- *ITU-T Rec. J.343.2* (2014), “Hybrid-NR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of non-encrypted bitstream data”
- *ITU-T Rec. J.343.3* (2014), “Hybrid-RRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a reduced reference signal and encrypted bitstream data”
- *ITU-T Rec. J.343.4* (2014), “Hybrid-RR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a reduced reference signal and non-encrypted bitstream data”
- *ITU-T Rec. J.343.5* (2014), “Hybrid-FRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a full reference signal and encrypted bitstream data”
- *ITU-T Rec. J.343.6* (2014), “Hybrid-FR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a full reference signal and non-encrypted bitstream data”

Actualmente el VQEG tiene en curso los siguientes proyectos²:

- Audiovisual HD (AVHD)

²Proyectos en curso, VQEG, [En línea], <http://www.its.bldrdoc.gov/vqeg/projects-home.aspx>

Capítulo 2. Estado del Arte en Evaluación de Calidad de Video

- HDR/WGG (High Dynamic Range Video / Wide Color 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 línea ³. 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 [57, 58], entre los cuales se incluyen el efecto de bloques, la borrosidad, el congelamiento de imágenes, efecto de entrelazado, entre otros. Para cada indicador se pueden establecer umbrales en lo que hace a la percepción de la degradación. De esta manera se pueden conformar modelos del tipo NR y obtener así indicadores de calidad, específicamente asociados a algunos de los tipos de degradaciones que normalmente se pueden encontrar en las imágenes y el sonido en transmisiones digitales.

En cuanto a los sistemas de televisión en agosto de 2012, ITU ha estandarizado la Recomendación ITU-R BT.2026, [59] 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 [60] y ITU-R BT.1908 [61], 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.

³www.GitHub.com/VIQET

2.5. Técnicas de desempeño y comparación de modelos

En lo que refiere a la evaluación del desempeño de los modelos objetivos usualmente se requieren ajustes no lineales para convertir el resultado de un modelo a una escala lineal con el MOS (1 a 5). Dichos procedimientos se describen en el informe del Grupo de Expertos en Calidad de Video (VQEG) aprobado en el año 2003 sobre la validación de modelos objetivos de calidad para la evaluación de la calidad de video [62]. Una vez realizado este ajuste, normalmente se utiliza el Coeficiente de Correlación Lineal de Pearson (PLCC) y la Raíz del Error Cuadrático Medio (RMSE), entre el MOS y el resultado de aplicar la transformación a los valores objetivos, a partir de estos indicadores es que se compara el desempeño de los modelos propuestos.

A continuación se define el PLCC y el RMSE para dos conjuntos de datos x e y :

$$PLCC = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (2.15)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (2.16)$$

Donde x e y representan los datos originales y su predicción, respectivamente. Por otro lado $\bar{x}$ e $\bar{y}$ son los valores medios de cada conjunto de datos. En el caso del PLCC el rango de variación es entre -1 y 1, si la relación entre los valores de x e y es perfectamente lineal y creciente (decreciente) entonces el resultado será de 1 (-1), en caso contrario si no hay correlación entre ambos conjunto de datos será de 0.

Adicionalmente se utilizan los Coeficientes de Correlación de Spearman (SROCC) y Kendall (KROCC) con el objetivo de medir monotonidad entre el resultado del modelo objetivo, sin aplicar el ajuste no lineal, y el MOS. Por ejemplo, ambos coeficientes varían entre -1 y 1, siendo 1 cuando existe una perfecta relación monótona creciente entre los conjuntos de datos y -1 cuando es monótona decreciente. Esto permite tener una noción del comportamiento de los modelos previo al cambio de escala.

Por otro lado, con el fin es evaluar si existe una diferencia estadística significativa en el desempeño de distintos modelos evaluados sobre una misma base de datos resulta necesario realizar un análisis estadístico. En este trabajo con el objetivo de realizar este análisis se siguen los criterios adoptados por VQEG en [46]. En primer lugar se utiliza la Transformación z de Fisher para convertir la PLCC a una variable z con distribución normal de acuerdo con la ecuación 2.17.

Capítulo 2. Estado del Arte en Evaluación de Calidad de Video

$$z = \frac{1}{2} \ln \left(\frac{1 + PLCC}{1 - PLCC} \right) \quad (2.17)$$

El test estadístico asume la hipótesis nula (H_0) de que no hay diferencia significativa entre los coeficientes de correlación de Pearson (PLCC). A continuación se calcula la variable Z_{jk} según la ecuación 2.18.

$$Z_{jk} = \frac{z_j - z_k}{\sqrt{\frac{2}{N-3}}} \quad (2.18)$$

Donde z_j es la Transformación z de Fisher del modelo j y z_k del modelo k y N es la cantidad de muestras que coincide la cantidad de clips de video utilizada para calcular la PLCC. Si $N > 30$, se puede asumir una distribución normal, en consecuencia Z_{jk} determina el nivel de confianza. Por ejemplo, si $Z_{jk} > 1,96$ implica que el modelo j es estadísticamente mejor que el modelo k con un nivel de confianza mayor al 95 %.

Capítulo 3

Televisión Digital Terrestre: norma ISDB-Tb

3.1. Introducción

Actualmente existen 4 estándares ampliamente difundidos de Televisión Digital Terrestre Abierta, éstos se mencionan a continuación:

- **ATSC**, Advanced Television Standards Committee
- **DVB-T**, Digital Video Broadcasting — Terrestrial
- **DTMB**, Digital Terrestrial Multimedia Broadcast
- **ISDB-T**, Integrated Services for Digital Broadcasting, Terrestrial

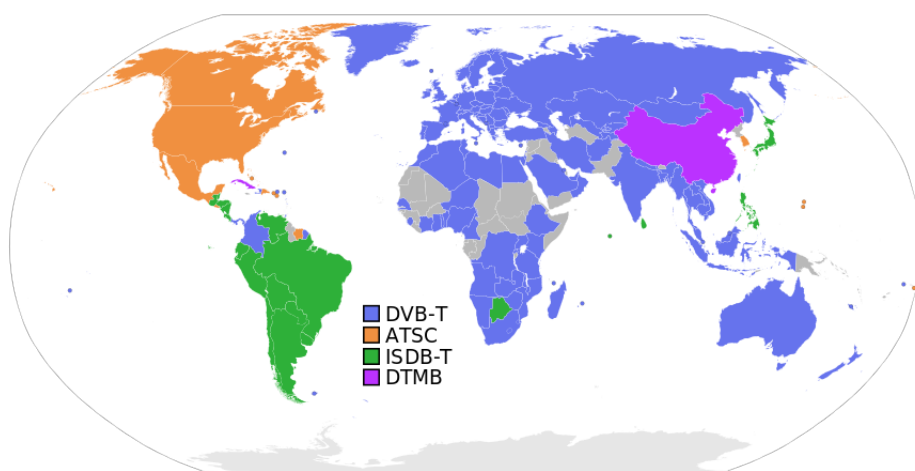


Figura 3.1: Ilustración de la utilización de los distintos sistemas de Televisión Digital por país.
Fuente: *Wikipedia*

Capítulo 3. Televisión Digital Terrestre: norma ISDB-Tb

Hoy en día la mayoría de los países desarrollados han adoptado la TV Digital terrestre por medio de alguno de estos estándares, mientras que en los países en vías de desarrollo la TVD se va abriendo camino. En la figura 3.1, a modo de ilustración general, se puede observar la adopción de estos estándares por país en la actualidad (2016). Uruguay, así como el resto de los países en la región, ha adoptado una versión modificada del estándar ISDB-T original. Ésta, según se detalla con claridad por Pisciotta et al. en [63] tiene su origen en los lineamientos para la TVD que el Gobierno Brasileño definió en noviembre de 2003, al poner en marcha el Sistema Brasileño de Televisión Digital Terrestre (SBTVD-T). Bajo este marco se estableció una alianza con el gobierno Japonés que permitió introducir algunas modificaciones al estándar ISDB-T, para adecuarlo a ciertas necesidades planteadas por Brasil. En virtud de dichas modificaciones, surgió el estándar ISDB-Tb, también conocido como ISDB-T Internacional, que fue adaptado por Brasil en junio de 2006. Uno de los acuerdos celebrados entre ambos países estableció la necesidad de difundir el estándar en toda la región, con el objetivo de formar un conglomerado que permitiera disminuir los costos de implementación y generara nuevas industrias relacionadas con la televisión. En Brasil, el organismo de estandarización que definió esta versión del estándar es la Asociación Brasileña de Normas Técnicas (ABNT) y sus normas son objeto de estudio y análisis en esta sección. El sistema de transmisión y recepción ISDB-T se encuentra enteramente descrito en los documentos ABNT NBR 15601 [64] y ABNT NBR 15604 [65] respectivamente. Sin embargo cabe destacar que la comunidad internacional ISDB-T ha llegado a acuerdos que se describen en el documento de armonización en [66], siendo éste tomado como referencia normativa por el decreto 143/013 del Poder Ejecutivo del gobierno uruguayo que establece las especificaciones técnicas a cumplir por los receptores de TVD en Uruguay. En esta sección y en el resto del trabajo, como previamente se ha mencionado, se refiere por las siglas ISDB-T al estándar adoptado por Uruguay y no a su versión original, que entre otras diferencias no utiliza compresión de video H.264 como si se utiliza en la versión brasileña.

En las siguientes secciones se repasan los conceptos y aspectos clave del sistema ISDB-T con respecto a la codificación fuente (codificación de video H.264), multiplexación (audio, video, datos), servicios (HD, SD, one_seg), codificación de canal, modulación y transmisión. No se pretende dar una mirada exhaustiva al estándar, sino dar una panorámica global y mencionar aquellos conceptos claves que se utilizarán en el presente trabajo.

3.2. Codificación de fuente

En una emisión de una estación de TVD bajo el sistema ISDB-T se pueden encontrar señales de video en resolución HD, SD, incluso señales específicas para el servicio móvil (dispositivos móviles) denominada one_seg, siendo su decodificación opcional en receptores fijos (denominados como *full_seg*). Estas señales están acompañadas de uno o varios audios, datos, aplicaciones interactivas (*Ginga*), subtítulos, entre otros. Este sistema permite emitir múltiples programaciones

3.2. Codificación de fuente

o servicios como suelen denominarse, con su respectivo video, audio/os y datos dentro de un canal físico de 6 MHz, con distintas grillas de contenidos. Se pueden combinar por ejemplo una señal en HD con dos señales en resolución estándar SD. Para ello el sistema ISDB-T utiliza compresión de video H.264/AVC conforme a la Recomendación ITU-T H.264 [67] e ISO/IEC 14496-10 [68]. Las especificaciones respecto de la codificación de video en alta definición, resolución estándar y resolución reducida se encuentran definidas en la norma ABNT NBR 15602-1 [69]. Por su parte para la codificación de audio el sistema utiliza compresión AAC (Advanced Audio Codec) con multiplexación y sincronización LATM/LOAS conforme a la normativa ISO/IEC 14496-3 [70]. En la norma ABNT NBR 15602-2 [71] se especifican los parámetros para estas señales de audio y el sistema de codificación y decodificación utilizado por el sistema. Por último con el fin de empaquetar y multiplexar los distintos servicios (señales de audio, video y datos) se utiliza el flujo de transporte MPEG-2 TS (*Transport Stream*).

En lo que sigue se hace especial foco en los principios de la compresión de video, los distintos formatos de pixel utilizados y los conceptos fundamentales utilizados en H.264. Por último se describe la obtención del flujo de transporte (*transport stream*) y la estructura de los paquetes TS.

3.2.1. Codificación de video

Principios de compresión de video

La compresión de video en un sistema de comunicaciones implica una pareja codificador/decodificador (en *COder/DECoder*), lo que usualmente se denomina como *CODEC* de video. En el caso de las señales de video existe una alta correlación espacial entre pixeles en un mismo cuadro y redundancia temporal entre cuadros en una secuencia. Estas redundancias son utilizadas por el codificador con el objetivo de disminuir la cantidad de información respecto de la secuencia original para ser guardada y/o posteriormente transmitida. Por su parte el decodificador convierte la señal codificada en una representación de la señal original. Si el proceso de codificación no produce pérdida de información a este se le denomina sin pérdida (*lossless*). En cambio, si la señal decodificada no es idéntica a la señal original, a este proceso se le denomina con pérdida (*lossy*) y puede derivar en una degradación de la calidad percibida de la señal de video (como es el caso de H.264). Los procesos de codificación de video con pérdida deben reducir la cantidad de bits necesaria para representar el contenido original y al mismo tiempo mantener la mayor fidelidad posible con este. Esto desencadena un conflicto dado que una reducción de bitrate usualmente deriva en una degradación de calidad de imagen, a este compromiso se le denomina como *Rate-Distortion* o tasa de distorsión en el campo de la Teoría de la Información y es el problema que intentan optimizar los codificadores de video con pérdida.

En la figura 3.2 se presenta un diagrama de codificador típico, este consta de tres componentes principales, un modelo temporal (estimación y compensación de movimiento), un modelo espacial (transformación y cuantización) y un codificador de entropía (*entropy coding*). En la entrada al codificador se introduce la

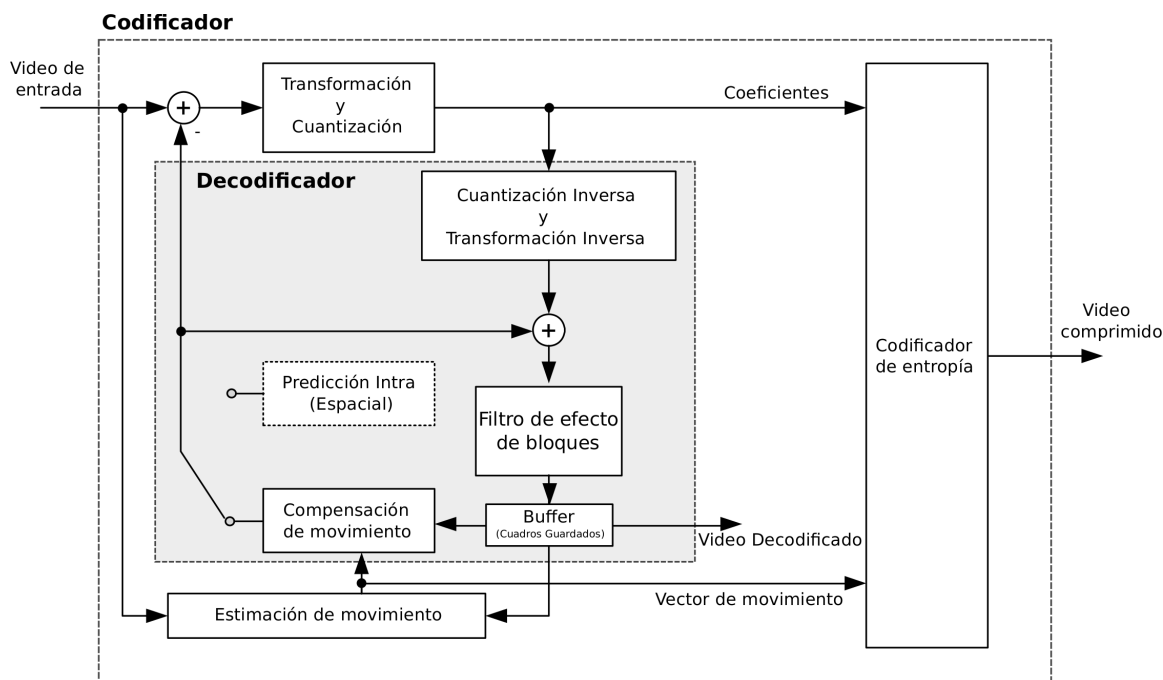


Figura 3.2: Diagrama de bloques de un codificador de video

secuencia de cuadros de video descomprimida. El modelo temporal se encarga de reducir la redundancia temporal entre cuadros en la secuencia. Para ello genera una predicción del cuadro actual, en H.264 esta predicción se crea a partir de cuadros anteriores y posteriores cargados en el *buffer* utilizando compensación de movimiento en base a bloques (modo de predicción *inter-frame*). Con esta predicción se genera un cuadro residual al realizar la diferencia entre el cuadro actual original y su predicción, además arroja los vectores de movimiento obtenidos en el proceso de estimación de movimiento. Este cuadro residual es la entrada al modelo espacial, encargado de reducir la redundancia espacial entre píxeles de un mismo cuadro (modo de predicción *intra-frame*). Esto se logra mediante un proceso de transformación y cuantización. El proceso de transformación realiza un cambio de dominio donde los valores de los píxeles se convierten en coeficientes, luego estos coeficientes son cuantizados y los valores insignificantes son removidos. Los coeficientes restantes proveen una representación comprimida del cuadro residual. En tanto a la salida del modelo espacial se arrojan estos coeficientes cuantizados. El proceso de cuantización en H.264 es realizado por un cuantizador escalar donde se define el parámetro QP (*quantization parameter*). Este es un índice que toma valores entre 0 y 51 y tiene asociado un escalón de cuantización denominado como *Qstep*. Respecto al proceso de transformación, H.264 utiliza la DCT (*Discrete Cosine Transform*) sobre bloques de tamaño 4×4 , ésta es integrada al proceso de cuantización para reducir el número de multiplicaciones dado que el sistema de codificación debe ser utilizado en transmisiones en vivo. Una vez que se obtienen los vectores de movimiento y los coeficientes cuantizados éstos se comprimen por medio de un codificador de entropía. Éste reduce la redundancia estadística en la

3.2. Codificación de fuente

información. Por ejemplo, en el caso de ISDB-T para la señal de video del servicio móvil (one_seg) se utiliza CAVLC (*Context-Based Adaptive Variable Length Coding*). En general estos codificadores de entropía asignan códigos de bits más largos o más cortos dependiendo de la probabilidad de ocurrencia de los eventos, con el objetivo de comprimir la información original. Por último cabe destacar que al guardarse los cuadros en el *buffer* los coeficientes se re-escalan (proceso inverso a la cuantización) y se les aplica la transformada inversa (IDCT), luego se le suma la predicción correspondiente y el bloque queda decodificado para su utilización en la predicción y compensación de movimiento. Se realiza de esta manera para que tanto el codificador como el decodificador contengan la misma información respecto a los cuadros de referencia para su uso.

Por su parte el decodificador, presentado en la figura 3.3 reconstruye la secuencia de cuadros a partir del bitstream comprimido. En primer lugar aplicando el decodificador de entropía obtiene los coeficientes cuantizados y los vectores de movimiento. Luego se reconstruyen los cuadros residuales y por medio de los vectores de movimiento y cuadros previamente decodificados se reconstruye la predicción del cuadro actual y se efectúa la suma con el correspondiente cuadro residual para obtener el cuadro actual decodificado.

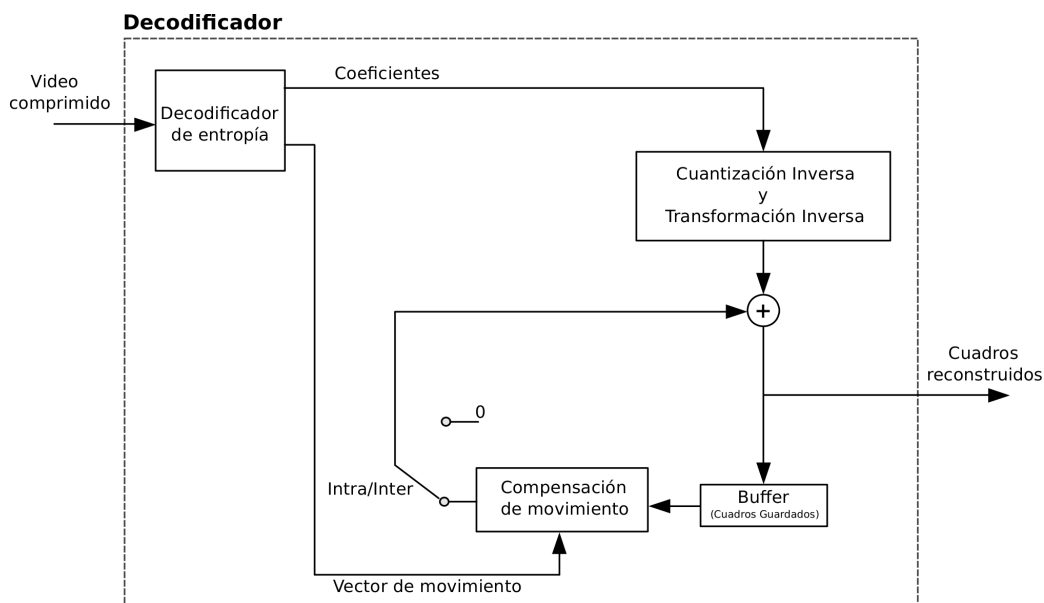


Figura 3.3: Diagrama de bloques de un decodificador de video

Formato de pixel

Con el fin de representar la información de color de un pixel se pueden utilizar varios espacios de colores. En el espacio RGB (*Red Green Blue*) una imagen de color es representada por sus componentes en rojo, verde y azul, siendo estos los colores primarios. Cualquier color puede ser creado al combinar distintas proporciones de rojo, verde y azul de manera aditiva. Por otro lado se encuentra el sistema $YCbCr$,

Capítulo 3. Televisión Digital Terrestre: norma ISDB-Tb

creado en base a la premisa de que el sistema visual humano (HVS) es menos sensible al color y más sensible la luminancia. En el sistema RGB cada uno de los colores es considerado con la misma importancia y la información es guardada con la misma resolución para cada color (por ejemplo, con 8 bits por color), sin embargo en el sistema $YCbCr$ la información de luminancia es separada de la de color. De esta manera se puede representar una imagen de manera más eficiente asignando una mayor resolución a la luminancia y una menor a la información de color. A este sistema generalmente se le denomina como YUV, donde como puede verse en la ecuación 3.1 Y se calcula como una combinación lineal de las señales R , G y B .

$$Y = k_r R + k_g G + k_b B \quad (3.1)$$

Mientras que la información de color se representa como una diferencia de color (*croma*) entre las respectivas señales R , G y B con la señal Y . A estas señales Cr , Cg , y Cb se les denomina señales de crominancia.

$$\begin{aligned} Cb &= B - Y \\ Cr &= R - Y \\ Cg &= G - Y \end{aligned} \quad (3.2)$$

Las componentes de croma que se conservan en el sistema $YCbCr$ como puede deducirse por su nombre son las componentes Y , Cb y Cr , extraídas de la señal RGB. La ventaja de este espacio de colores es que se puede disminuir la resolución de las señales Cb y Cr sin incidir notoriamente en la calidad visual de la imagen. Esto permite realizar una compresión de la imagen reduciendo la cantidad de información a ser guardada y el ancho de banda para su posterior transmisión. Luego para ser desplegada en pantalla la imagen debe ser reconvertida a su formato RGB. Para efectuar el pasaje entre estos sistemas existen normalizaciones, por ejemplo la recomendación ITU-R BT.601 [72] establece los parámetros de codificación para esta transformación entre espacios de color con aplicación en televisión digital. Los parámetros utilizados en el sistema bajo estudio se encuentran definidos en la norma ABNT NBR 15602-1 [69].

Por otro lado, en la figura 3.4 se ilustran distintos patrones de muestreo soportados por H.264/AVC. Se puede ver que en el caso del patrón YUV444 por cada pixel se tiene una muestra de la señal Y y una de cada señal Cr y Cb respectivamente, mientras que para el patrón YUV422 (usualmente denominado YUY2) esto sucede cada 2 muestras de Y , por último para el patrón YUV420 cada 4 muestras de Y correspondientemente.

Se puede decir que el formato YUV420 o también conocido como YUV420p es el más popular, probablemente debido a la alta compresión lograda en las señales de croma. En la figura 3.5 se muestra la posición de cada una de las muestras Y , U y V que conforman un cuadro de video con formato YUV420. Cuando comienza el cuadro en el bitstream de la secuencia de video se colocan los componentes asociados a Y , luego los asociados a la componente U y por último los asociados a la componente V según se muestra en la figura.

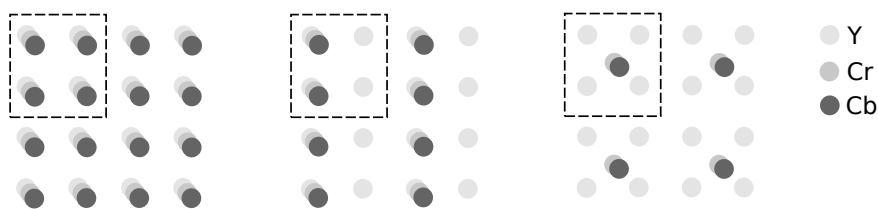


Figura 3.4: Patrones de muestreo YUV444,YUV422,YUV420 (progresivos)

Cuadro YUV420



Posición en el Bitstream YUV

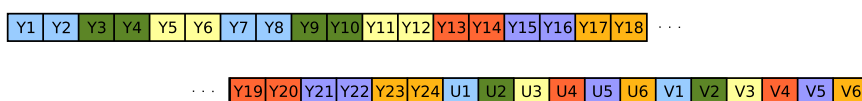


Figura 3.5: Cuadro YUV420p. Ilustración basada en una imagen extraída de Wikipedia.

A modo de ejemplo en caso de que se tenga una señal PAL de 720×576 píxeles en formato YUV420, se tendrán 720×576 muestras de Y, además de 360×288 muestras Cr y Cb respectivamente. Si para cada píxel se toman 8 bits se obtienen entonces $720 \times 576 \times 8 + 2 \times (360 \times 288 \times 8) = 4976640$ bits. Por otro lado si se utiliza YUV444, se tendrá $720 \times 576 \times 8 \times 3 = 9953280$ bits, siendo el doble de tamaño.

ISDB-T permite barrido progresivo y entrelazado. En el modo entrelazado un cuadro de una secuencia se conforma por un campo par y un campo impar. En el caso del barrido progresivo se trata del cuadro completo. La ventaja del barrido entrelazado frente al progresivo es que manteniendo el número de cuadros por segundo el barrido entrelazado es capaz de generar el doble de campos por segundo dándole una sensación de mayor continuidad a la escena. Por ejemplo en el caso de PAL los campos se generan a una tasa de 50 Hz. Cuando el contenido es reproducido se tiene una sensación de movimiento con mayor continuidad que si fueran cuadros progresivos a una tasa de cuadro de 25 Hz. En caso de barrido entrelazado, los patrones de muestreo así como el esquema de codificación se adaptan a los campos par e impar del cuadro respectivamente.

3.2.2. H.264/AVC

H.264, el estándar de compresión de video utilizado por ISDB-T, ha mostrado su superioridad respecto a sus predecesores como H.263 logrando tasas de compresión más altas. Este sistema utiliza un esquema de estructura que hace uso de compensación de movimiento en base a bloques, cuantización, transformación y codificación de entropía. En común con los estándares de codificación anteriores, H.264 no define explícitamente una pareja Codificador/Decodificador, sino que define la sintaxis de un flujo de bits codificado en conjunto con el método de decodificación de este flujo de bits.

En MPEG-2 las secuencias de cuadros de videos son agrupadas en “*Group of Pictures*” (GOP). Cada una de estas secuencias codificadas contiene cuadros de tipo I (Intra), P (Predictive) y B (Bidirectional). Los cuadros I son codificados únicamente por un modelo espacial (Intra) y son utilizados como referencia para la predicción de cuadros P y B. Los cuadros P son codificados utilizando información de cuadros I y P previos en base a estimación y compensación de movimiento basada en bloques. Mientras los cuadros tipo B son codificados en base a información de cuadros previos y futuros. El tamaño del GOP usualmente se mide en función de la distancia en cuadros entre cuadros I consecutivos. Por su parte H.264/AVC extiende el funcionamiento de su antecesor, donde cada cuadro codificado consiste de un número determinado de macro bloques (MB), por ejemplo de 16×16 muestras de la señal de luma (Y) y 8×8 muestras de las señales de *Cr* y *Cb*. A su vez cada cuadro de video es codificado en una o varias tirillas (*slices*). Cada una de éstas puede contener desde un MB hasta el número total de MB en el cuadro entero. En este último caso un slice abarcaría todo un cuadro (1 slice por cuadro). Cada cuadro puede estar compuesto por varios tipos de slices, por ejemplo de tipo I, B o P. Los *slices* de tipo I (Predicción-*Intra*) contienen MB de tipo I. En este caso cada bloque o macrobloque dentro del slice es predicho con la información incluida en el resto de los macrobloques decodificados en el mismo *slice*. Respecto de los *slice* de tipo P (Predicted) estos contienen MB de tipo P. Cada uno de estos son predichos a partir de un cuadro de referencia. Por otra parte los slices de tipo B contienen MB de tipo B que se predicen a partir de más de un cuadro de referencia, por ejemplo con cuadros anteriores y posteriores. En caso de los MB de predicción-*Inter*, es el caso de los tipo P, estos pueden ser subdivididos en sub-bloques de 16×16 , 8×16 , 16×8 muestras de luma y su muestras *Cr*, *Cb* asociadas. Además estas particiones pueden subdividirse en porciones menores de 8×8 , 8×4 , 4×8 y 4×4 muestras de luma (Y) y las muestras asociadas de *Cb* y *Cr*.

El sistema ISDB-T utiliza un conjunto de perfiles y niveles definidos por H.264, entre ellos se encuentra el perfil *High*, *Main* y *Baseline*. Cada uno de estos soporta un conjunto específico de funcionalidades. Por ejemplo el perfil *Baseline* (utilizado para la señal *one_seg*) utiliza solo *I-Slices* y *P-Slices*. Mientras el perfil *Main* y *High* incluyen codificación de video entrelazado, *B-Slices* y CABAC (*Context-Based Arithmetic Coding*) como codificador de entropía. Por otro lado se establece un conjunto de niveles para cada perfil que limitan, por ejemplo la resolución de video, requerimientos de memoria, tasa de procesamiento, etc.

3.2. Codificación de fuente

A continuación se presenta, a modo de ejemplo, un posible esquema de codificación para una señal de video en definición estándar y su audio asociado. Por más ejemplos referente a esquemas de codificación para señales de video y audio se recomienda al lector a referirse a [73].

- Señal SD:
 - Video:
 - Estándar de codificación: ISO/IEC 14496-10 (MPEG-4 AVC/H.264)¹
 - Resolución: $720 \times 576i$ @ 50 Hz
 - Relación de aspecto: 4:3
 - Perfil y nivel: high@L4.0
 - Bitrate: 3.8 Mbps, constante
 - Audio:
 - Estándar de codificación: MPEG-4 AAC
 - Transporte y multiplexación: LATM/LOAS
 - Perfil: HE-AAC
 - Frecuencia de muestreo: 48 kHz
 - Bitrate: 128 kbps
 - Número de canales: estéreo (2/0)
 - Cuantización: 16 bits

3.2.3. Generación del flujo de transporte MPEG-2 TS

La generación de contenido en una estación de TV se produce por medio de distintos equipos como cámaras, selectores de video, etc. Estos entregan señales o flujos elementales de alta velocidad que son distribuidos para luego ser comprimidos para su posterior transmisión. En el caso del video digital generalmente se utiliza la interfaz SDI (*Serial Digital Interface*). En la figura 3.6 se ilustra un diagrama de bloques de un *encoder* genérico. Éste es capaz de adquirir estas señales de audio y video generadas dentro de un estudio de TV y a la salida entregar el TS asociado. Para ello el *encoder* comprime las señales según lo especificado en las secciones anteriores, con esto se obtienen los *elementary streams* (ES) de cada una de éstas. La información contenida en cada uno de estos *elementary streams* a su vez es segmentada en paquetes denominados PES (*Packetized Elementary Streams*) que luego son nuevamente divididos en paquetes de menor extensión denominados paquetes TS (*Transport Stream Packets*). Cada uno de estos flujos se multiplexa para formar el TS del programa, que contiene los paquetes que transportan la información asociada a un servicio. Una transmisión de TVD puede contener varios servicios, por lo tanto los *transport streams* obtenidos para cada servicio se deben nuevamente multiplexar para conformar el denominado MPEG2-TS, flujo

¹El término MPEG-4 AVC/H.264 hace referencia al estándar de codificación de video H.264/AVC utilizado por el sistema bajo estudio.

Capítulo 3. Televisión Digital Terrestre: norma ISDB-Tb

multiprogramación con todos los programas que se desean transmitir incluyendo datos y tablas para la correcta recepción de la señal de TVD en los receptores. Cabe destacar que un servicio generalmente incluye un video en resolución HD o SD acompañado de uno o varios audios, datos, subtítulos, etc. En tanto, se pueden combinar varios servicios en SD o efectuar una combinación HD/SD en una misma transmisión. Esto es siempre que el bitrate total del TS resultante no supere la capacidad máxima del sistema. Esta capacidad se puede configurar dependiendo de los parámetros de transmisión seleccionados, que se describen en la próxima sección.

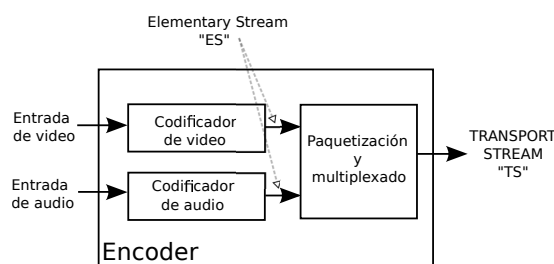


Figura 3.6: Diagrama en bloques de un *encoder* genérico

Los paquetes TS tienen una longitud fija de 188 Bytes. Como se puede observar en la figura 3.7, constan de un encabezado (*head*) de 4 Bytes y una carga útil (*payload*) de 184 Bytes. Adicionalmente al encabezado fijo, la carga útil puede incluir un campo de adaptación que extiende al encabezado manteniendo la longitud de los paquetes constante. A continuación se describen cada uno de los elementos del encabezado:

- ***sync***: Byte de sincronización con valor 47_H
- ***transport_error_indicator***: Indicador de existencia de errores ocasionados durante el proceso de transmisión en el paquete, si el bit se pone en 1, el decodificador descarta el paquete.
- ***payload_unit_start_indicator***: Si este bit se encuentra en 1, indica que la información contenida en la carga útil representa el inicio de un nuevo paquete PES o una tabla de información (por ejemplo, PSI). En el caso de que se transporte una tabla el primer byte del *payload* es un indicador que apunta al comienzo de la primer sección de la tabla mientras que si es un PES este coincide con su inicio.
- ***transport_priority***: Indicador de nivel de prioridad de 1 bit de información, si este se pone en 1 indica que el paquete es de mayor prioridad que aquellos que tienen el mismo PID y el bit en 0.
- ***PID (Packet Identifier)***: Identifica el tipo de dato que transporta el paquete TS, sea un video, audio o datos. Estos identificadores tendrán valores distintos para cada tipo de dato. Debe tenerse en cuenta que existen valores

3.2. Codificación de fuente

reservados por el sistema, según se ilustra en la tabla 3.1. Por más información respecto a los criterios de uso de PID referirse a la norma ABNT NBR 15603-1 [74].

Tabla 3.1: Asignación de valores PID

Valor	Descripción
0000 _H	PAT (<i>Program Association Table</i>)
0001 _H	CAT (<i>Conditional Access Table</i>)
0002 _H a 000F _H	Reservado
0010 _H	NIT (<i>Network Information Table</i>)
0011 _H a 1FFE _H	Otras tablas y datos, excluyéndose PAT, CAT, NIT y paquetes nulos
1FFF _H	Paquetes nulos

- ***transport_scrambling_control***: Indica si el paquete se encuentra codificado (criptografía).
- ***adaptation_field_control***: Indica la presencia o no de los campos de adaptación en la carga útil, es decir indica que solamente sea carga útil sin campo de adaptación o solamente sea campo de adaptación sin carga útil o por último sea campo de adaptación seguido de carga útil.
- ***continuity_counter***: Es un contador progresivo de paquetes identificados con el mismo PID, tiene 4 bits y el conteo es secuencial de 00_H a FF_H, una vez que llega a su fin se reinicia. Los paquetes nulos no se cuentan.
- ***adaptation_field***: Campo de adaptación, opcional.

Respecto del campo de adaptación, este muchas veces cumple con la función de mantener constante el largo de 188 Bytes del paquete. A continuación se describe la información más relevante incluida en este.

- ***adaptation_field_length***: Largo en Bytes del campo de adaptación, sin incluir el largo de 1 Byte de este campo. El campo de adaptación puede tener hasta 184 Bytes de longitud en caso de que no contenga carga útil.
- ***discontinuity_indicator***: Indicador de discontinuidad, por ejemplo en la base de tiempo en relación a los PCR o en el contador de continuidad. Si se pone en 1, existe discontinuidad.
- ***PCR (Program Clock Reference)***: Es la referencia del reloj principal utilizado en los procesos de decodificación.

Por más información referente de las banderas incluidas en la estructura del paquete TS previamente descrito se aconseja al lector a referirse [63].

Cuando esta secuencia de paquetes TS llega al decodificador éste se debe sincronizar con el Byte de sincronización (*sync*) con valor 47_H. Dado que este no es un valor prohibido se debe detectar una sucesión de cinco bytes de sincronismo

Capítulo 3. Televisión Digital Terrestre: norma ISDB-Tb

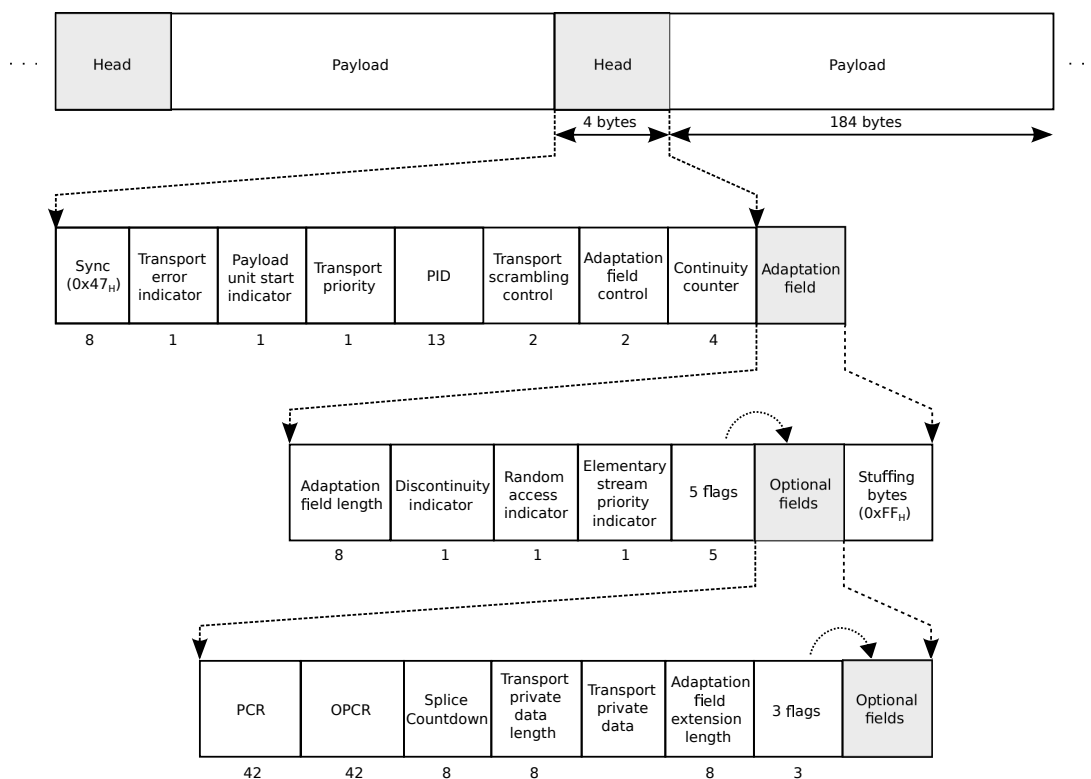


Figura 3.7: Estructura del encabezado de un paquetes TS

consecutivos válidos para entrar en sincronismo. Una vez que esto sucede se interpreta el encabezado y se busca el PID para detectar los distintos flujos. Los paquetes TS pueden contener partes de paquetes PES de audio o video, secciones parciales o completas de tablas definidas por el sistema, referencias de sincronización PCR de flujos PES o paquetes nulos. En lo que refiere a las tablas de datos, estas son estructuras que se utilizan para enviar ciertos tipos de datos al receptor, están compuestas por varias secciones y como previamente se mencionó cada una se encuentran identificada por un PID especificado por el sistema. A continuación se describen brevemente las tablas PAT, CAT y PMT que conforman la tabla PSI (Program Specific Information).

- **PAT** (*Program Association Table*) Ésta es la primer tabla que busca el receptor una vez que éste entra en sincronismo, por lo tanto es enviada periódicamente. Contiene información acerca del contenido general del flujo y contiene el listado de programas e información de la red emisora. Es una tabla de asociación que permite identificar los paquetes que contienen la PMT de cada programa.
- **PMT** (*Program Map Table*) El PID es asignado por la tabla PAT. Cabe destacar que el sistema reserva el PID 1FC8_H para la tabla PMT utilizada en el servicio móvil one_seg. Contiene la información específica de un programa, con sus respectivos PID de video, audio, datos y PCR. Habrá tantas tablas

3.2. Codificación de fuente

PMT como programas emitidos. Dentro de esta tabla se incluyen descriptores con información de los parámetros de codificación de video y audio, entre otros. Especialmente se destaca el descriptor *stream_type* que informa sobre las características de los flujos elementales de los servicios transportados. Por ejemplo, para el caso del flujo elemental de video este debe tomar el valor de $1B_H$. Este valor indica que la codificación de video este conforme a la Recomendación ITU H.264 e ISO/IEC 14496-10. En el caso del flujo elemental de audio este debe tomar el valor 11_H , este valor indica que la codificación de audio este conforme a la norma ISO/IEC 14496-3. Por más información acerca de los descriptores referirse a la norma ABNT NBR 15602-3 [75].

- **CAT** (*Conditional Access Table*) Se utiliza para el acceso condicional (encriptación) y suministra los PID de la tabla EMM (*Entitlement Management Message*).

Por otro lado se tiene la tabla SI (*System Information*), conjunto formado por las siguientes tablas, en su mayoría opcionales, que transportan información disponible para servicios o eventos: BAT (*Bouquet Association Table*), NIT (*Network Information Table*), SDT (*Service Description Table*), EIT (*Event Information table*), TDT (*Time Data Table*), TOT (*Time Offset Table*), ST (*Stuffing Table*), etc. Dentro de éstas se destaca la tabla **NIT** (*Network Information Table*). Ésta debe enviarse de manera obligatoria y contiene información sobre la cantidad de flujos TS disponibles, multiplexores, frecuencia de transmisión, número de canal de la red, servicios, denominaciones y sus números de programa, sistema de transmisión empleado por la red y datos técnicos del transmisión. Además los descriptores que posee llevan la siguiente información: nombre de la red, lista de servicios, sistema de distribución terrestre, información de emergencia, recepción parcial, etc.

Con el fin de concluir la sección, en la figura 3.8 se muestra una captura de pantalla de un encoder profesional configurado según los parámetros de codificación fuente y flujo de transporte MPEG-2 TS previamente explicados. Se puede observar que toma el video directamente desde una entrada HDMI mientras que para el audio utiliza estéreo (2 canales). Para la compresión de video se utiliza el *codec* H.264 con perfil *High*, bitrate constante (CBR, *Constant BitRate*) a 8 Mbps con un largo de GOP de 30 cuadros y una resolución de salida de 1920×1080 con barrido entrelazado. En cuanto al audio se utiliza el definido por el sistema a un *bitrate* de 128 Kbps. La salida es un flujo MPEG-2 TS que se extrae por una interfaz ASI y por red (Ethernet). Por último se agregan los PIDs de Audio, Video y PCR de manera manual. Cada una de las componentes transmitidas en el flujo TS tiene un bitrate dado y la suma de todas ellas da como resultado el bitrate resultante del TS generado. En el caso del ejemplo el encoder pide de manera manual fijar el bitrate del TS de salida, en tanto se debe asegurar que este sea superior a la suma de todas las tasas individuales fijándose en 9Mbps.

Capítulo 3. Televisión Digital Terrestre: norma ISDB-Tb

Video Source	HDMI	Output Interface	ASI+Ethernet(CBR)
Audio Source	Channels 1+2	Output Format	MPEG-2 TS
Audio Codec	SBTV-D LATM	Dest IP Address	192.168.0.200
Audio Bit Rate	128 kbits	Dest Port	4322
Video Resolution	1080i (1920x1080)	TS Bitrate	9000k
Rate Control	CBR	ASI PCR Interval	50
Video Bit Rate	8000k	PID Audio,Video,PCR	274 273 256
Video Profile	High	Local IP Setup	
Gop Size	30 Frames	Local IP Address	192.168.0.102
Video burst size	300ms	Local Netmask	255.255.255.0
Video Codec	H.264	Default Gateway	192.168.0.1

Figura 3.8: Captura de pantalla de Encoder utilizado para la generación de un flujo MPEG-2 TS ISDB-T. Equipo: Z3 Technology, LLC MVE-02.

3.3. Codificación de canal

ISDB-T utiliza modulación OFDM sobre un canal de $6MHz$. A su vez el canal se divide en 13 subcanales llamados “segmentos”. Cada uno de estos tiene un ancho de banda $6/14MHz \approx 428,57KHz$, el ancho de banda restante se distribuye a modo de guarda a cada uno de los lados de la señal modulada en el espectro radioeléctrico. En la figura 3.9 se presenta el diagrama en bloques del sistema de transmisión. Éste incluye el proceso de codificación de canal. Los primeros tres bloques representan al Remultiplexor, Código Corrector Reed-Solomon (Codificación externa) y división en capas jerárquicas, respectivamente.

Una vez que los *transport streams* pasan por el Remultiplexor se obtiene el denominado BTS (*Broadcast Transport Stream*). Este flujo tiene un bitrate fijo de aproximadamente $32,5Mbps$. En el remultiplexor a cada uno de los paquetes de 188 bytes se le agrega 16 bytes de información, parte de ésta refiere a su respectiva capa jerárquica y se utiliza posteriormente para direccionar cada uno de los paquetes en el proceso de división en niveles jerárquicos. Por otro lado, previo a la división se aplica el proceso de corrección externa o *outer code*. Éste es el encargado de aplicar el algoritmo Reed-Solomon, según se denota en el bloque del diagrama por “RS(204,188)”, añadiendo 16 bytes de redundancia en lugar de los 16 bytes de información agregados en el BTS original. De esta manera el sistema es capaz de corregir hasta 8 bytes de información. El sistema ISDB-T define una estructura en capas jerárquicas: “A”, “B”, “C”, cada una de éstas porta diferentes contenidos multimedia y puede ser configurada independientemente, por ejemplo con distintos códigos convolucionales, esquemas de modulación y entrelazado en tiempo. El código convolucional o *inner code* es el segundo algoritmo de corrección utilizado. Éste se puede configurar con los siguientes *ratios*: $1/2$, $3/4$, $5/6$, $7/8$. Por ejemplo, si se utiliza $7/8$ cada 7 bits el código agrega 1 bit de redundancia. Cada una de las

3.3. Codificación de canal

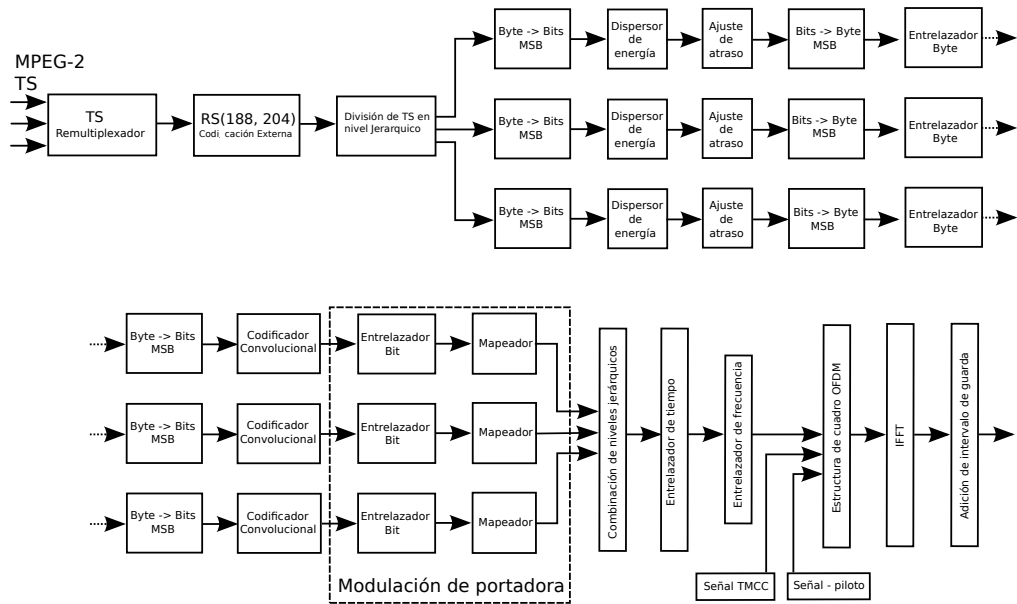


Figura 3.9: Diagrama en bloques del sistema de transmisión

capas puede utilizar un *ratio* distinto permitiendo un compromiso entre bitrate útil e inmunidad contra errores. Respecto de los esquemas de modulación el sistema permite utilizar los siguientes: DQPSK, QPSK (donde cada portadora transmite 2 bits de información por símbolo), 16-QAM y 64-QAM, donde cada portadora transmite 4 y 6 bits por símbolo respectivamente. Por otro lado el entrelazado en tiempo o *time interleaving* reorganiza los símbolos transmitidos con el objetivo de fortalecer la señal contra ráfagas de errores. Los parámetros del *time interleaver* se pueden configurar como $\{0, 4, 8, 16\}$, $\{0, 2, 4, 8\}$ ó $\{0, 1, 2, 4\}$ dependiendo del modo seleccionado 1, 2 o 3 respectivamente. Estos modos provistos por el sistema ISDB-T mantienen el bitrate constante y aumentan (modo 3) o disminuyen (modo 1) el número total de portadoras utilizadas por segmento. Particularmente el modo 1, 2 y 3 utilizan 108, 216 y 432 portadoras por segmento con un largo efectivo por símbolo de $252\mu s$, $504\mu s$, $1004\mu s$. Como muchos sistemas OFDM, ISDB-T además agrega un prefijo cíclico a cada símbolo OFDM. Con esto logra inmunizar la señal contra interferencia entre símbolos debido al efecto multicamino o *multi-path* en el canal de radiofrecuencia. Este prefijo es una copia de la última parte del símbolo OFDM, y se antepone a este, haciendo la señal periódica [76]. A este prefijo cíclico se le denomina “intervalo de guarda”. El sistema ofrece cuatro posibles valores: $1/4$, $1/8$, $1/16$ y $1/32$, siendo estos su longitud expresada como una fracción de la longitud de símbolo activo. Cuanto mayor el intervalo de guarda mayor inmunidad contra el *multi-path* [77], sin embargo el bitrate útil disminuye dado que el tiempo por cada símbolo OFDM aumenta.

En la figura 3.10 se presenta la tasa de datos o bitrate resultante para un único segmento y en la figura 3.11 los mismos parámetros se especifican para los 13 segmentos. Se puede observar como la tasa total de datos durante la transmisión jerárquica varía dependiendo de los parámetros de configuración. El volumen

Capítulo 3. Televisión Digital Terrestre: norma ISDB-Tb

Modulación de la portadora	Código convolucional	Número de TSP transmitidos por cuadro	Tasa de datos ^a kbps			
			Intervalo de guarda 1/4	Intervalo de guarda 1/8	Intervalo de guarda 1/16	Intervalo de guarda 1/32
DQPSK	1/2	12/24/48	280,85	312,06	330,42	340,43
	2/3	16/32/64	374,47	416,08	440,56	453,91
	3/4	18/36/72	421,28	468,09	495,63	510,65
QPSK	5/6	20/40/80	468,09	520,10	550,70	567,39
	7/8	21/42/84	491,50	546,11	578,23	595,76
16QAM	1/2	24/48/96	561,71	624,13	660,84	680,87
	2/3	32/64/128	748,95	832,17	881,12	907,82
	3/4	36/72/144	842,57	936,19	991,26	1021,30
	5/6	40/80/160	936,19	1 040,21	1 101,40	1 134,78
	7/8	42/84/1 68	983,00	1 092,22	1 156,47	1 191,52
64QAM	1/2	36/72/144	842,57	936,19	991,26	1 021,30
	2/3	48/96/192	1 123,43	1 248,26	1 321,68	1 361,74
	3/4	54/108/216	1 263,86	1 404,29	1 486,90	1 531,95
	5/6	60/120/240	1 404,29	1 560,32	1 652,11	1 702,17
	7/8	63/126/252	1 474,50	1 638,34	1 734,71	1 787,28

^a Esa tasa de datos representa la tasa de datos (bits) por segmento para parámetros de transmisión:
tasa de datos (bits) = TSP transmitidos x 188 (bytes/TSP) x 8 (bits/byte) x 1/longitud del cuadro.

Figura 3.10: Tasa de datos de un único segmento. Tabla basada en la norma ABNT NBR 15601 *Televisión Digital Terrestre - Sistema de Transmisión*

transmitido por los 13 segmentos es igual a la suma de todos los volúmenes de datos transmitidos por cada uno de los segmentos, que puede ser determinado de acuerdo con la figura 3.10. Por ejemplo, se puede observar manteniendo constante el esquema de modulación y el código convolucional como incrementa la tasa de datos conforme disminuye el intervalo de guarda. Como también, al fijar el intervalo de guarda, el esquema de modulación y disminuir el *ratio* del código convolucional aumenta la tasa de datos. Por otro lado se puede observar que al enviar mas información (bits) por símbolo en los esquemas de modulación y mantener constante el intervalo de guarda y el *ratio* del código convolucional aumenta la tasa de datos, por ejemplo al pasar de 16-QAM a 64-QAM, de 4 a 6 bits/símbolo, respectivamente.

A diferencia de los sistemas DVB-T y ATSC, ISDB-T permite transmitir una señal de baja resolución, señal *one_seg*. Esta señal se emite en el segmento central del espectro. A este segmento se lo denomina como “segmento 0” y se define en la capa “A” usualmente con parámetros de transmisión robustos, como por ejemplo modulación QPSK. Los receptores capaces de sintonizar esta señal usualmente son de tipo móvil como teléfonos celulares o tabletas y se les denomina receptores *one_seg*. Por otro lado a los receptores capaces de sintonizar y demodular la información del resto de los segmentos que se encuentran, por ejemplo en este caso, en las capas “B” y “C”, se les denominan como *full_seg*.

A modo de ejemplo, en la figura 3.12 se presenta una captura de pantalla de la configuración de los parámetros de codificación de canal y de transmisión en un

3.3. Codificación de canal

Modulación de la portadora	Código convolucional	Número de TSP transmitidos (Modos 1/ 2/ 3)	Tasa de datos Mbps			
			Intervalo de guarda 1/4	Intervalo de guarda 1/8	Intervalo de guarda 1/16	Intervalo de guarda 1/32
DQPSK	1/2	156/312/624	3,651	4,056	4,295	4,425
	2/3	208/416/832	4,868	5,409	5,727	5,900
	3/4	234/468/936	5,476	6,085	6,443	6,638
	5/6	260/520/1040	6,085	6,761	7,159	7,376
	7/8	273/546/1092	6,389	7,099	7,517	7,744
16QAM	1/2	312/624/1248	7,302	8,113	8,590	8,851
	2/3	416/832/1664	9,736	10,818	11,454	11,801
	3/4	468/936/1872	10,953	12,170	12,886	13,276
	5/6	520/1040/2080	12,170	13,522	14,318	14,752
	7/8	546/1092/2184	12,779	14,198	15,034	15,489
64QAM	1/2	468/936/1872	10,953	12,170	12,886	13,276
	2/3	624/1248/2496	14,604	16,227	17,181	17,702
	3/4	702/1404/2808	16,430	18,255	19,329	19,915
	5/6	780/1560/3120	18,255	20,284	21,477	22,128
	7/8	819/1638/3276	19,168	21,298	22,551	23,234

Figura 3.11: Tasa total de datos para 13 segmentos. Tabla basada en la norma ABNT NBR 15601 *Televisión Digital Terrestre - Sistema de Transmisión*.

equipo modulador y transmisor profesional de TVD.

Se puede observar que se configura el canal 21 de la banda UHF asociado a la frecuencia 515,143MHz para ser utilizado como canal físico. Además se utiliza el modo 3 con un intervalo de guarda de 1/32 y se activa la recepción parcial para dispositivos *one_seg*. Respecto de la división en capas jerárquicas, en la capa “A” se selecciona solo 1 segmento (Segmento “0”), se utiliza modulación QPSK, código convolucional de 2/3 y la duración del entrelazado en tiempo en 2. Con estos valores se obtiene un bitrate útil de 453,910kbps para la capa “A”. Luego se colocan los restantes segmentos en la capa “B”, dejando la capa “C” vacía. En ella se utiliza modulación 64-QAM, código convolucional de 7/8 y duración del entrelazado en tiempo de 4. De esta manera se obtiene un bitrate útil de 21,45Mbps. Siendo el bitrate total útil obtenido de 21,9Mbps.

Respecto del contenido que se desea emitir, siguiendo con el ejemplo anterior, en la capa “A” se puede asociar una señal *one_seg* de resolución reducida. Cabe destacar que para el correcto funcionamiento del sistema el bitrate total asociado al TS que se desea emitir debe ser inferior al bitrate útil obtenido al realizar la

Capítulo 3. Televisión Digital Terrestre: norma ISDB-Tb

Transport Stream: LabTVD

Configuration | DVB-C | DVB-T | DVB-S | ISDB-T

Broadcast type: 13-segment TV Broadcast

Frequency: Channel 21 515142857 Hz

Area Code: 1

Virtual Channel (Remote Control Key ID): 8

TS Name: LabTVD 1

Transmission Mode: Mode 3: 8k

Guard Interval: 1/32

Partial Reception: ☒ Enabled partial reception in the Layer A (implies in Layer A with only 1 segment)

Emergency: ☐ Enabled emergency flag

Hierarchical Layers

	Layer A	Layer B	Layer C
Segments	1	12	0
Modulation	QPSK	64-QAM	DQPSK
Convolutional Rate	2/3	7/8	5/6
Time Interleaving Length	2	4	8
Bitrate (bps)	453910	21447360	0
Total Bitrate (bps)	21901270		

Save

Figura 3.12: Configuración de parámetros de transmisión para una señal ISDB-T en un equipo profesional modulador de TVD. Equipo: EITV Payout Pro.

configuración de los parámetros de transmisión. En este caso el bitrate útil obtenido para la capa “A” es de 453,910Kbps. Si desea transmitir una señal *one_seg* que supere dicho bitrate no se obtendrá el resultado deseado. Por otro lado en la capa “B”, se tiene un bitrate útil de 21,45Mbps, aquí se puede rellenar el espacio con distintas variantes, por ejemplo con 2 señales en resolución HD, o una señal HD y su espejo en SD. Cada una de estas señales puede tener múltiples audios, subtítulos, *closed caption* y datos. Además como en el caso anterior la suma del *bitrate* individual de estas señales no debe superar el *bitrate* útil disponible.

Por más ejemplos de configuración de parámetros de transmisión se recomienda al lector referirse a [73]. En este documento se especifican distintas configuraciones que pueden ser utilizadas para transmisión con el objetivo de ensayar receptores *full_seg* de tipo Televisor o STB (*Set Top Box*) respecto de la correcta recepción y presentación de la señal.

Capítulo 4

Evaluación Subjetiva de la calidad percibida de video en TV Digital

4.1. Introducción

En esta sección se presentan los resultados obtenidos tras la recolección de 8900 calificaciones en evaluaciones subjetivas de calidad percibida de video con aplicación en Televisión Digital Abierta. Se describe el diseño, desarrollo y metodología de trabajo utilizada para su realización y se analizan los resultados obtenidos. Cabe destacar que la base de datos se encuentra a disposición para la comunidad de investigación en QoE, con el fin de entrenar, verifica y validar modelos objetivos de evaluación de calidad de video [2] [78].

En lo que sigue se muestran los resultados obtenidos, se incluye una discusión sobre los criterios para la selección de los clips de video, los parámetros de codificación, los tipos de perturbaciones, incluyendo los distintos patrones de pérdida de paquetes utilizados, así como la experiencia recogida durante las sesiones de evaluación. Además, se describe brevemente la aplicación informática utilizada para automatizar las pruebas, disponible para la comunidad de QoE [2] [79]. 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. Estas pruebas fueron realizadas en el marco del proyecto VQI [2].

4.2. Descripción general de la base de datos

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). 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

Capítulo 4. Evaluación Subjetiva de la calidad percibida de video en TV Digital

con degradaciones controladas y fue concebido para entrenar y verificar modelos objetivos. Mientras que el segundo 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 [5] 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 [7]. Cada video clip se presentó de manera independiente al evaluador, y conforme con la recomendación 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.

4.3. Selección y preparación del material

A continuación se describen los conjuntos *Set I* y *Set II*.

Set I

Con el fin de ser utilizado para entrenar y verificar modelos de estimación de calidad de video, un conjunto de clips de video de prueba debe cubrir un amplio rango de situaciones en las que los modelos serán aplicados. Como criterio de selección de los videos, en general se considera 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* [14] y de la base de datos *IRCCyN IVC1080i Video Quality Database* [80]. Con el objetivo de obtener una medida de movimiento y de complejidad de las secuencias se utilizaron dos indicadores, denominados MV y SAD. El primero es una estimación del promedio de amplitudes de los vectores de movimiento generados durante el proceso de compresión y el segundo es el promedio de la suma de diferencias absolutas de los bloques residuales computados a lo largo del clip de video. El cómputo de estos indicadores, MV y SAD, se describe con mayor detalle en la sección 5.1.1.

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* [81]. La tabla 4.1 muestra una vista previa representativa de cada una de las secuencias originales seleccionadas, incluyendo el tipo de contenido, SAD y MV.

Los parámetros de codificación utilizados se describen en la tabla 4.2 ¹. Para obtener 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

¹Con el objetivo de desplegar los videos en pantalla se decidió codificar los videos utilizando barrido progresivo.

4.3. Selección y preparación del material

de 25 clips de video en alta resolución degradados por artefactos de codificación. Con el ánimo de incluir errores en transmisión, un grupo reducido de los videos obtenidos 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 [82].

Tabla 4.1: Secuencias originales seleccionadas para el *Set I HD* & *Set SD*

Vista previa	Descripción
	Nombre: Fox & Bird Descripción: Dibujos animados de un zorro y un pájaro Contenido: Dibujos animados SAD: 90 MV: 16
	Nombre: Golf Descripción: Juego de Golf Contenido: Deportes SAD: 156 MV: 2
	Nombre: Concierto Descripción: Jean-Michel habla en un concierto Contenido: Música SAD: 273 MV: 11,6



Nombre: Fútbol

Descripción: Gol durante un partido de fútbol Alemán

Contenido: Deportes

SAD: 283

MV: 14



Nombre: Voile

Descripción: Un barco se mueve en el agua.

Contenido: Movie

SAD: 198

MV: 8,3

Tabla 4.2: 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 el análisis de las muestras grabadas se encontraron períodos con pérdidas homogéneas (posiblemente debido a desvanecimiento de la señal) con un largo promedio de más de 10 segundos y períodos con ráfagas de errores (muchas pérdidas en un período corto de tiempo). En la mayoría de los casos el largo de las ráfagas no superó el segundo y fue seguido de períodos de varios segundos sin pérdidas. Cerca de la mitad de las pérdidas consistieron de paquetes TS individuales, esto se puede ver analizando la bandera denominada *Continuity Counter* en el cabezal del paquete TS asociado al PID (identificador) del *stream* de video. Basados en estas observaciones distintos patrones de pérdidas de paquetes TS fueron simulados y aplicados a cada uno de los video clips de acuerdo a la tabla 4.3. En la figura 4.1

4.3. Selección y preparación del material

se muestran dos vistas previas por cada secuencia de origen de cuadros afectados por pérdida de paquetes.



Figura 4.1: Vista previa de cuadros afectados por simulación de errores en transmisión en cada uno de los contenidos seleccionados

Para obtener el *Set I SD* cada uno de los cinco secuencias originales en alta resolución (HD) fue convertida a resolución estándar (SD) utilizando la herramienta *FFmpeg*. Para ello por defecto se utilizó interpolación bicubica. Análogamente, estas secuencias fueron codificadas con compresión H.264/AVC nuevamente uti-

Capítulo 4. Evaluación Subjetiva de la calidad percibida de video en TV Digital

Tabla 4.3: 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

lizando *FFmpeg* a cuatro tasas de bits distintas. Se escogió un rango de valores normalmente utilizado para secuencias en definición estándar en TVD: 0.7, 1.5, 2.8, 4.0 Mbps, los parámetros de codificación utilizados se muestran en la tabla 4.2. Con este procedimiento se obtienen 20 clips de video degradados con artefactos de codificación.

Al igual que en el *Set I HD*, a un grupo reducido de estos clips de video se le aplicó un procedimiento de extracción de paquetes TS individuales. Los patrones de pérdida de paquetes utilizados se muestran en la tabla 4.3.

Como resultado se obtuvo en total un centenar de video clips en alta resolución y un centenar en resolución estándar.

Set II

El *Set II* fue concebido como un conjunto de clips de video independientes para realizar validación de modelos, al igual que el *Set I* se compone de un centenar 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)
- IBBBPBBBPBBI (largo GOP = 12, hasta 2 B seguidos)
- IPPPPPPPPPI (largo GOP = 12, hasta 0 B seguidos)

Además se notó que la cantidad de *slices* por cuadro era variable según el canal. Se registraron dos casos:

- 1 slice por cuadro

4.3. Selección y preparación del material

- 6 slices por cuadro

La figura 4.2 muestra una vista previa de algunos de los clips utilizados sin degradaciones por errores en transmisión. Por otro lado en la figura 4.3 se muestran algunas de las secuencias afectadas por pérdida de paquetes TS tal como fueron registradas de la transmisión de la TV abierta.



Figura 4.2: Vista previa de cuadros sin pérdida de paquetes TS de grabaciones tomadas por aire



Figura 4.3: Vista previa de cuadros afectados por errores en transmisión en algunas secuencias grabadas de aire

4.4. Metodología de evaluación

Se montó un ambiente controlado de acuerdo a la recomendación ITU-R BT.500. Una imagen de éste se puede ver en la figura 4.4. Se ubicaron cinco asientos en la sala respetando el ángulo de visión y la distancia a la pantalla. Las condiciones de visión se pueden ver en la tabla 4.4. Como pantalla se utilizó un televisor de 42" con tecnología LED.



Figura 4.4: Ambiente controlado de pruebas subjetivas

En la tabla 4.5 se muestra para cada conjunto la cantidad de sesiones que requirió, el total de participantes, la variabilidad de participantes por sesión, el total de calificaciones adquiridas por el sistema, la cantidad de participantes que fallaron el test de visión o daltonismo (Ishihara), los participantes identificados como *outliers* y por último la cantidad de usuarios con los cuales se calculó la opinión media (MOS). La identificación de *outliers* se realizó mediante la correlación de Pearson (PLCC) entre las calificaciones individuales por participante y el promedio de calificaciones. Si la correlación no supera un umbral definido se considera al participante como atípico, para esto se utilizó un umbral de 0,70.

En cada sesión se utilizó un centenar de clips de video, y en promedio cada una de éstas duró unos cincuenta minutos, nunca excediendo los 60 minutos. El tiempo de sesión fue medido desde el momento en que los espectadores se encontraban sentados y leyendo las instrucciones hasta que terminaban de calificar el último video. Fue utilizado un máximo de cinco espectadores por sesión, sentados en la misma sala, calificando el mismo clip de video a la vez. Uno se podría preguntar si la presencia de más de un espectador puede alterar o distraer las tareas de calificación. Del comienzo, se notó necesaria la presencia de una figura de autoridad por lo que cada sesión contó con una persona a cargo dentro de la sala. Como precaución adicional antes de comenzar las pruebas se le indicó a los usuarios que era una tarea individual con el objetivo de evadir que los usuarios miraran las calificaciones de los demás, manteniendo así la independencia entre calificaciones. Sobre un total de 27 sesiones solo una fue descartada debido a factores de distracción entre los sujetos de prueba que derivaron en una notoria baja correlación en comparación con el promedio de calificaciones. Como información complementaria, solo una sesión fue interrumpida dado que uno de los usuarios tuvo que retirarse por razones personales.

Finalmente, cabe destacar que los resultados obtenidos en todos los conjuntos

4.4. Metodología de evaluación

Tabla 4.4: Condiciones generales de observación para efectuar evaluaciones subjetivas en el entorno de laboratorio

Parámetros	ITU-R BT.500-13	Medidas en el Laboratorio
Relación entre la luminancia de pantalla inactiva y el valor de pico de la luminancia:	$\leq 0,02$	0,011
Relación entre la luminancia de la pantalla, cuando sólo se muestra el nivel del negro en una sala completamente oscura, y la correspondiente al blanco más intenso:	$\approx 0,01$	0,013
Brillo y contraste de la imagen:	Establecido vía PLUGE (véanse las Recomendaciones ITU-R BT.814 y ITU-R BT.815)	Establecido vía PLUGE
Ángulo máximo de observación con respecto al normal (este valor se aplica a las pantallas de tubo de rayos catódicos (TRC), para otro tipo de pantallas se están estudiando los valores adecuados)	30°	28°
Relación entre la luminancia de fondo detrás del receptor de imágenes y el valor de cresta de luminancia de la imagen:	$\approx 0,15$	0,15
Cromaticidad del fondo:	D_{65}	Blanco
Otra iluminación de la sala	Débil	Débil

(*Set I HD*, *Set I SD*, *Set II HD* y *Set II SD*), son considerados formales según la recomendación ITU-R BT.500 dado que ésta establece un mínimo de 15 evaluadores para considerar la evidencia como formal.

Capítulo 4. Evaluación Subjetiva de la calidad percibida de video en TV Digital

Tabla 4.5: Sesiones de pruebas subjetivas

Características	SetIHD	SetISD	SetIIHD	SetIISD	Total
Cantidad de sesiones	6	6	8	7	27
Total de participantes	25	19	27	18	89
Participantes por sesión	4 a 5	3 a 5	2 a 5	1 a 5	1 a 5
Total de calificaciones	2500	1900	2700	1800	8900
Participantes que fallaron test de visión o daltonismo (Ishihara)	4	0	3	0	7
Post-screening (outliers)	1	3	0	2	6
Participantes remanentes	20	16	24	16	76

4.5. Sistema de automatización de pruebas subjetivas

El sistema informático utilizado para la realización de las pruebas subjetivas se presenta en [79] y se encuentra disponible en [2]. Se diseñó y desarrolló en el marco del Proyecto VQI. El sistema utiliza dispositivos móviles (teléfonos inteligentes o tabletas) que permiten a los evaluadores enviar las calificaciones de cada uno de los clips de video a un servidor web. Los teléfonos inteligentes hoy en día son utilizados por muchas personas, no siendo difíciles de conseguir. Con la utilización de este sistema se consiguieron las siguientes ventajas:

- No hay necesidad de utilizar dispositivos electrónicos especiales
- Varios usuarios pueden participar de la misma sesión. El limite en cantidad de usuarios está dado por los valores asociados al angulo de visión, distancia y espacio del ambiente de pruebas establecido en las recomendaciones
- Fue diseñado para posibilitar la extensión en su uso a evaluación de calidad en escenarios reales (ej. crowdsourcing, personas en sus hogares etc.) no solo en ambientes de prueba controlados
- Pueden ser implementadas escalas continuas o discretas.
- No se utilizan formularios en papel ni lápiz, tampoco dispositivos conectados vía cable, haciéndolo más amigable para el usuario y para los investigadores que conducen las pruebas.

Además el sistema ofrece la posibilidad de crear y configurar distintas sesiones con videos de prueba, que luego se desplegarán automáticamente a los usuarios, facilitando la tarea de recolección y procesamiento de las calificaciones. Éstas son

4.6. Análisis de resultados subjetivos

ingresadas al sistema mediante teléfonos inteligentes en sincronización con la reproducción de los clips de video. Además en cada sesión los videos se despliegan de manera aleatoria según las recomendaciones. Finalmente se exportan los resultados de la sesión en un hoja de cálculo.

Cuando comienza una sesión cada evaluador debe registrarse y completar un pequeño formulario incluyendo: edad, sexo, nivel educativo, tipo de televisor que en general utiliza para mirar televisión y el número promedio de horas que mira televisión por semana. Una vez que todos los evaluadores de la sesión de prueba se registran y completan el formulario la sesión comienza con el primer video de prueba. Cuando finaliza cada video, el sistema pide a cada usuario que evalúe la calidad del video presentado en la aplicación móvil. Una vez que todos los usuarios evalúan el clip de video, el sistema despliega el siguiente video. Este proceso se repite hasta que los usuarios evalúan el último clip dando fin a la sesión.

4.6. Análisis de resultados subjetivos

Tras analizar los resultados del *Set I* fue concluido que no existe diferencia significativa entre las opiniones entre hombres y mujeres, tampoco en diferencia de edades. Sin embargo se observó cierta tendencia en como las personas que miran televisión en pantallas de tipo CRT tienden a calificar mejor la calidad de imagen (nótese que la pantalla utilizada fue un TV LED de 42 "). Esta tendencia se puede visualizar en la figura 4.5, publicada [78]. Además se observó cierta tendencia de la gente que mira más horas de televisión en dar peores calificaciones. Esto se puede apreciar en la figura 4.6 (publicada en [78]). Sobre el *Set II* se verificaron las mismas tendencias.

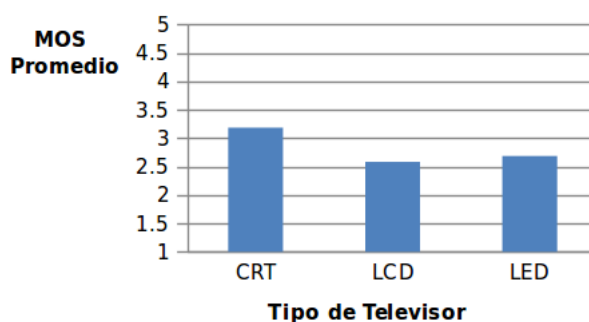


Figura 4.5: MOS promedio en función del tipo de Televisor (CRT, LCD o LED) que utilizan los sujetos de prueba

Por otro lado, en la figura 4.7 se muestran los resultados de calidad obtenidos para cada una de las secuencias originales procesadas sin pérdida de paquetes, pérdida uniforme, en ráfagas ("bursts"), con un porcentaje de 0,1 % o 10 % de pérdida dentro de cada burst, para distintos bitrates. Se puede observar que con pérdidas uniformes la calidad percibida se ve fuertemente afectada, mientras que, si no hay pérdidas la calidad en general tiende a disminuir con la caída del bitrate.

Capítulo 4. Evaluación Subjetiva de la calidad percibida de video en TV Digital

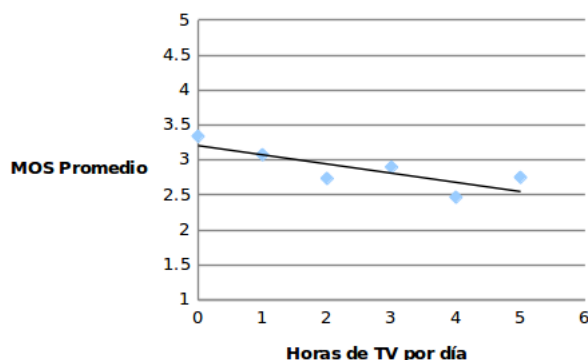


Figura 4.6: MOS promedio en función de la cantidad de horas que los usuarios miran televisión

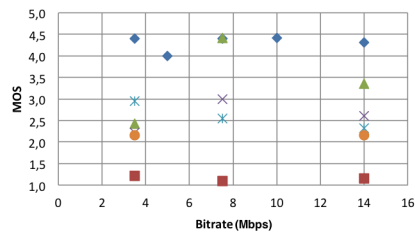
Por otro lado las ráfagas de errores tienen resultados variables. En algunos casos se puede ver que la calidad no se ve afectada y en otros disminuye de manera significativa. En el siguiente capítulo se explotan las características de estos clips de video y los resultados de las evaluaciones subjetivas (MOS) para conseguir estimar su calidad percibida por medio de un modelo paramétrico objetivo.

Hasta ahora se ha mostrado la metodología aplicada y el MOS obtenido tras la realización de evaluaciones subjetivas de calidad de video según lo establecido por las normalizaciones utilizadas. Sin embargo, cabe destacar que aunque el MOS sea comúnmente el indicador aceptado como medida de calidad este enfoque ignora la diversidad en la distribución de calificaciones de los observadores. Estas diferencias se dan debido a que el mismo estímulo (en este caso un clip de video) normalmente no recibe la misma calificación por todos los observadores del panel, lo cual es de esperar, ya que la noción de calidad de video es, justamente, subjetiva. Normalmente en los reportes de este tipo de evaluaciones que se pueden encontrar en la literatura esta diversidad de calificaciones usualmente se expresa en función de la desviación estándar y el intervalo de confianza para cada una de las opiniones medias (MOS) obtenidas. Para ello se utiliza el total de calificaciones individuales por clip de video y generalmente una confianza del 95 %.

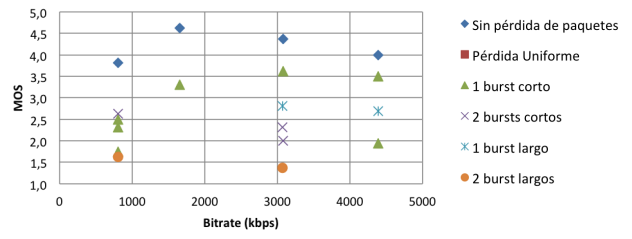
La mayoría de los trabajos existentes sobre la medición objetiva de calidad se centra sólo en la predicción de la opinión media sin tener asociada la estimación de un indicador de variabilidad. De esta manera se ignora la diversidad que subyace en el proceso de puntuación subjetiva, y como resultado, no permite un análisis más detallado sobre el grado de fiabilidad de la predicción objetiva en términos de variabilidad. Este hecho podría potencialmente disminuir la confianza en las estimaciones objetivas, particularmente cuando existe un gran desacuerdo entre las calificaciones para una secuencia (alta variabilidad).

Este enfoque asociado al estudio de la variabilidad de las calificaciones subjetivas no es nuevo. En la literatura distintos autores han discutido sobre la utilización del MOS como indicador único. Por ejemplo, en [83] se sugiere que al tomar en cuenta la diversidad de las calificaciones subjetivas recolectadas se incrementa la información extraída de la base de datos utilizada. Por otro lado en [84] los autores aplican este principio, donde en adición al tradicional MOS es estudiado un indica-

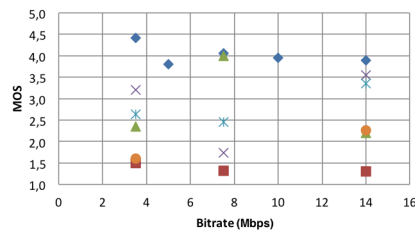
4.6. Análisis de resultados subjetivos



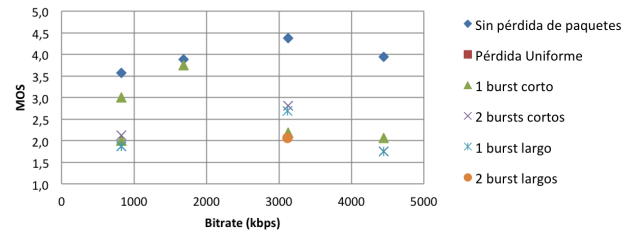
(a) Fox & Bird HD



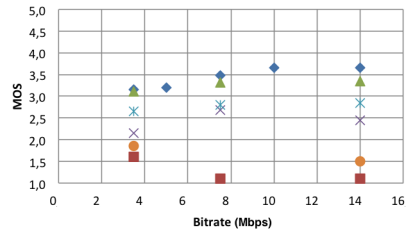
(b) Fox & Bird SD



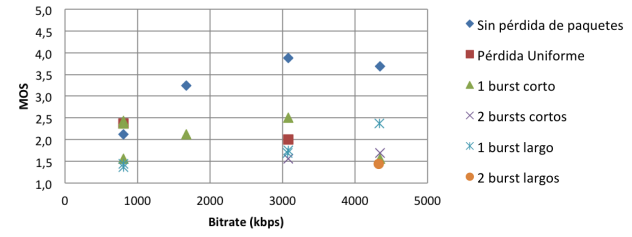
(d) Golf HD



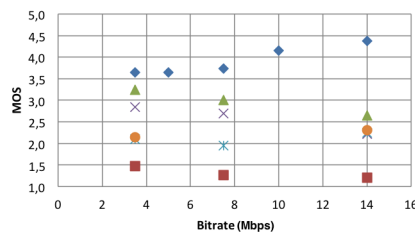
(e) Golf SD



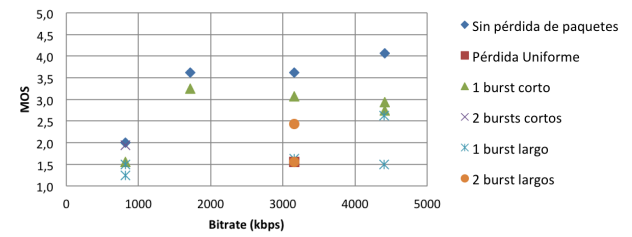
(g) Concierto HD



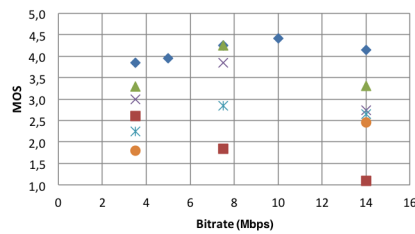
(h) Concierto SD



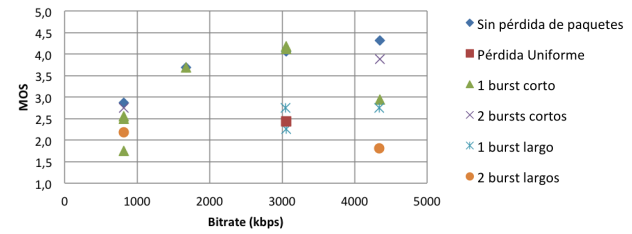
(j) Futbol HD



(k) Futbol SD



(m) Voile HD



(n) Voile SD

Figura 4.7: Ilustración de resultados de calidad obtenidos para el Set I HD & SD

Capítulo 4. Evaluación Subjetiva de la calidad percibida de video en TV Digital

dor asociado a la desviación estándar de las calificaciones denominado como *SOS* (*Standard deviation of Opinion Scores*). La relación matemática entre el MOS y *SOS* es definida y varios conjuntos de datos son analizados en función del MOS y este valor de *SOS*.

En [3], analizamos esta cuestión. En este trabajo se presenta un método de medida de la variabilidad de notas subjetivas y un modelo basado en aprendizaje automático (*machine learning*) para abordar esta problemática. Para demostrar la utilidad de la solución propuesta se utilizó el escenario práctico de la Televisión Digital Abierta por medio de la base de datos descrita en este capítulo. En lo que sigue se describen las principales ideas, reflexiones y resultados obtenidos que extienden el análisis realizado hasta ahora de los resultados subjetivos obtenidos.

Con el fin de ilustrar con un ejemplo estas ideas, en la figura 4.8 se presenta la relación entre el MOS y la desviación estándar para cada una de las secuencias contenidas en la base de datos generada.

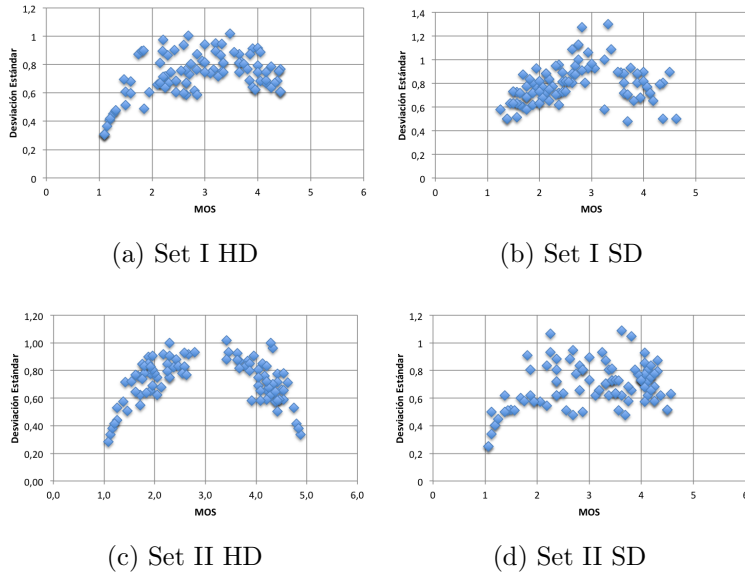


Figura 4.8: Desviación Estándar en función del MOS para los conjuntos Set I HD, Set I SD, Set II HD y Set II SD

Se puede observar que la dispersión asociada a esta relación tiene forma de arco donde para un MOS alto o bajo la desviación se hace menor y va aumentando a medida que el MOS llega al valor central de la escala utilizada (en este caso 3). Este tipo de tendencia se pueden encontrar en evaluaciones subjetivas de calidad aplicada a otros servicios de entrega de video. En este caso la escala utilizada para las calificaciones individuales va de 1 a 5. Si un video clip tiene una calidad mala, los observadores lo calificarán con números bajos con límite en 1. Caso contrario si el clip es de excelente calidad las notas saturan en 5. Por otra parte si el promedio de las calificaciones es 3, es donde puede existir mayor desacuerdo entre las calificaciones, lo cual deriva en una mayor incertidumbre.

4.6. Análisis de resultados subjetivos

Igualmente, cabe destacar que el hecho de utilizar la desviación estándar como suplemento del MOS no siempre se traduce en un resultado fácil de interpretar. Ésta, por ejemplo, no deriva en un conocimiento de la simetría de las calificaciones respecto de la media. En tanto sería deseable obtener un indicador que permita obtener una interpretación más directa de la diversidad de las opiniones y sobre todo que permita un mejor entendimiento sobre la aplicación o servicio que se esté brindando.

Un MOS bajo para un servicio determinado indica mala calidad y por tanto la decepción con el servicio por parte de sus usuarios. Pero incluso si el MOS es alto, solo con este valor no es posible saber cuántos usuarios están realmente satisfechos con el servicio y cuántos no lo están. Es sabido que los usuarios que no se encuentran satisfechos con un servicio son más propensos a compartir sus experiencias negativas que aquellos que sí están satisfechos. Por lo tanto, una experiencia negativa de un usuario podría derivar en un riesgo de avalancha, donde la experiencia negativa se extienda a otros clientes potenciales o actuales que van a ver el servicio de una manera más negativa que antes, sin a priori tener una mala experiencia. Como ya se ha destacado, la mayoría de los métodos objetivos simplemente ignoran esta diversidad de opiniones por parte de los usuarios, y en su lugar se centran sólo en las calificaciones medias.

Con el fin de superar este problema se estudiaron varios métodos para expresar la diversidad teniendo en cuenta el punto de vista del valor de negocio. Desde este punto de vista es útil tener información de los usuarios satisfechos e insatisfechos. Por ejemplo, desde la perspectiva de la calidad, los usuarios satisfechos no requieren gestión de la calidad de su servicio, aunque esto no quiere decir que los usuarios satisfechos no deben considerarse en absoluto en la comercialización del servicio en general. En tanto un posible indicador puede derivarse directamente de la respuesta a la pregunta “¿Cuántos usuarios están insatisfechos con el servicio?”. Siguiendo esta línea en [3] se introduce el indicador denominado como “Porcentaje de Usuarios Insatisfechos” (PDU, *Percentage of Dissatisfied Users*) como suplemento al MOS. Éste se define en la ecuación 4.1 e indica, dado un MOS, el porcentaje de usuarios que podrían dar una opinión inferior a cierto umbral previamente definido por el operador del servicio.

$$PDU = \frac{\#(OS < U)}{N} \times 100 \quad (4.1)$$

Donde OS (*opinion score*) representa las calificaciones individuales de un estímulo en particular, U es el umbral definido para la aplicación y N es el número de calificaciones. Con el objetivo de presentar este indicador aplicado a la base de datos generada, en la figura 4.9 se ilustra para todos los conjunto de datos (Set I HD, Set I SD, Set II HD, Set II SD) la dispersión del indicador PDU en función del MOS, donde se utiliza un umbral U con valor de 3.

En estas gráficas con U tomando el valor de 3, se puede observar una relación de tipo sigmoidal entre el indicador PDU y el MOS, incluso se podría pensar en un aproximación estadística simple donde se explote esta tendencia. Donde el PDU se estime por medio de una función de densidad de probabilidad *Gaussiana*, la cual

Capítulo 4. Evaluación Subjetiva de la calidad percibida de video en TV Digital

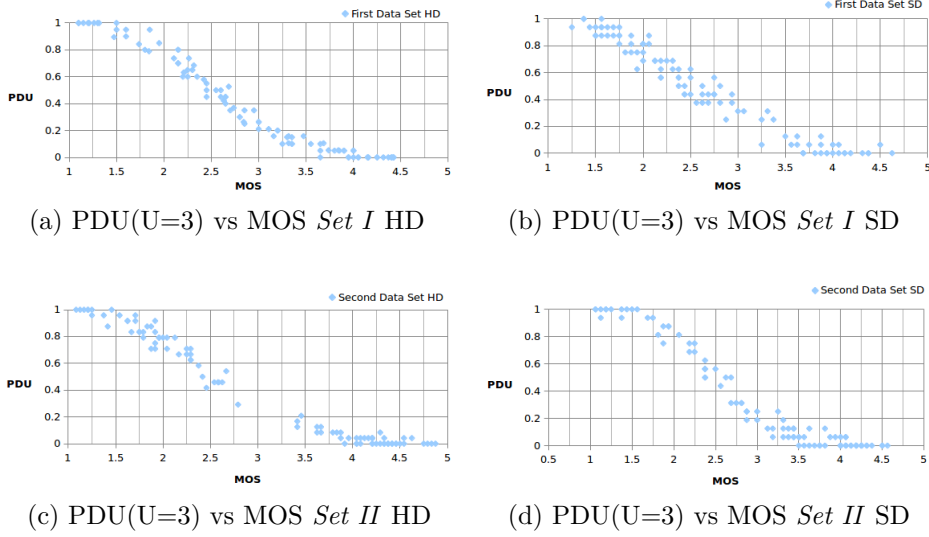


Figura 4.9: PDU vs MOS para los conjuntos Set I HD, Set I SD, Set II HD y Set II SD

se presenta en la ecuación 4.2.

$$\widehat{PDU}_i = P(1 \leq x \leq U) = \int_1^U \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} dx \quad (4.2)$$

Donde U es el umbral elegido, μ_i representa el MOS para el video i y σ_i es la desviación estándar de todas las calificaciones individuales asociadas al video i . Sin embargo, debido al hecho de que en un escenario real no es posible contar con los valores de μ_i y σ_i en un sistema de estimación de calidad en tiempo real no sería posible aplicar esta solución de manera directa.

Particularmente en [3] se brindan dos posibles soluciones con el fin de utilizar esta ecuación. Ambas, consisten de estimar el PDU a partir de la propia predicción del MOS (por medio de un modelo de calidad) y de estimar σ , lo cual se realiza de dos maneras distintas. La primera opción consiste de fijar un valor único de σ , al tomar el promedio de las desviaciones estándar de los videos de los conjuntos Set I HD y Set I SD. Mientras la segunda de tomar un valor variable de σ por medio del ajuste de una curva en los valores de σ_i en función del MOS para los conjuntos Set I HD y Set I SD.

En la figura 4.9 se ha utilizado un umbral de U con valor 3, pero en algunas aplicaciones comerciales este umbral U podría variar, por ejemplo, subir si se tratara de video a demanda (VoD, *Video on Demand*) donde los usuarios pagarán una cuota mensual o *Pay Per View*. Con el objetivo de evaluar la relación PDU-MOS en función del Umbral U en la figura 4.10 se ilustra la dispersión PDU contra el MOS para todos los conjuntos de videos. Donde U toma los valores: 2,3,4 y 5.

En cada una de estas gráficas no solo se ilustran los puntos sino también se delimita el espacio muestral (área sombreada) de todos las posibles combinaciones

4.6. Análisis de resultados subjetivos

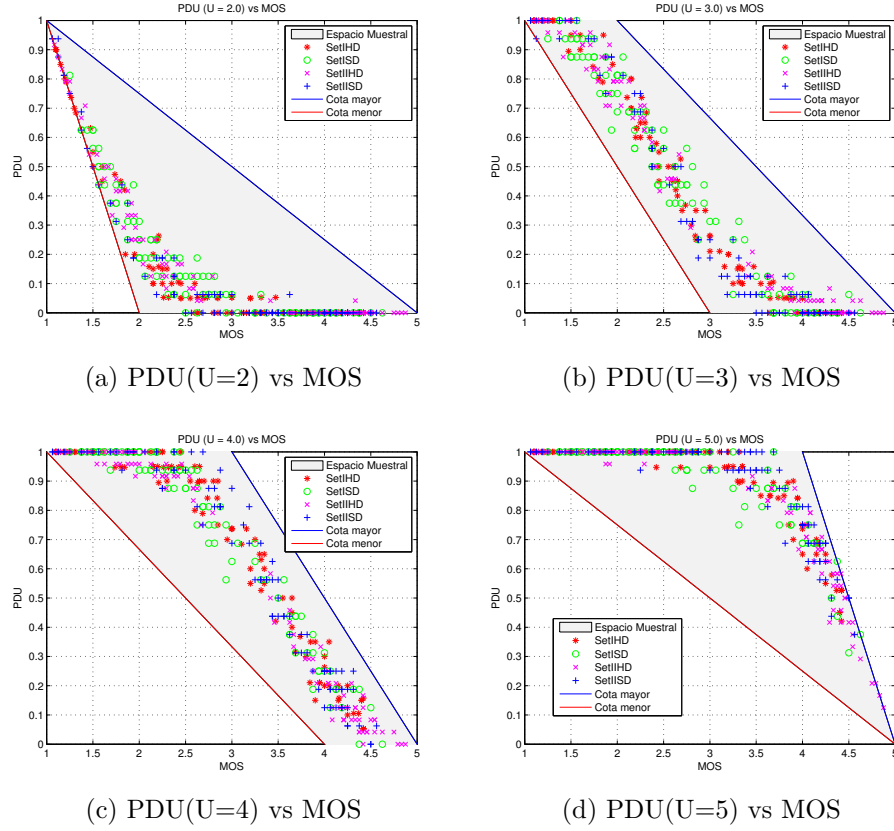


Figura 4.10: PDU vs MOS, se toma U con los siguientes valores 2,3,4 y 5 y se aplica para los conjuntos Set I HD, Set I SD, Set II HD y Set II SD

(PDU, MOS). Para tal fin se extendió el trabajo presentado en [3], donde se realizaron varias simulaciones. Para ello se definió un conjunto de vectores de largo n con cada elemento tomando un valor de 1,2,3,4 o 5 de manera aleatoria. Como una primera aproximación se tomaron 2^{10} vectores de largo 10 con distribución uniforme. Luego se fue definiendo el área en base a estos resultados prácticos tomando la envolvente convexa de los puntos. Por último se presentan las cotas máximas y mínimas que pueden tomar los valores de PDU en función del MOS.

Estas cotas se pueden expresar de manera paramétrica según las ecuaciones 4.3 y 4.4, respectivamente:

$$PDU_{cota_max} = \frac{5 - x}{6 - \lceil U \rceil} \quad (4.3)$$

$$PDU_{cota_min} = \frac{\lceil U \rceil - x}{\lceil U \rceil - 1} \quad (4.4)$$

Donde x toma valores en el intervalo real $\in [1, 5]$. Estas ecuaciones fueron formuladas en base a las gráficas presentadas y cada una de estas cumple con las cotas

Capítulo 4. Evaluación Subjetiva de la calidad percibida de video en TV Digital

impuestas por medio de las simulaciones. Es interesante observar estas cotas teóricas y sus aplicaciones, por ejemplo para $U = 3$ dado un valor de $MOS = 2$ se puede decir que el PDU será superior al 50 %, es decir más de la mitad de los usuarios estarán insatisfechos con el servicio. Por otro lado en caso contrario si $MOS = 3$ habrá un porcentaje no mayor al 66,67 % de usuarios insatisfechos mientras que teóricamente este valor puede bajar a 0 %. En tanto estas cotas además ofrecen un valor en sí mismas y podrían ser utilizadas para la toma de decisiones a nivel de negocio en un sistema de entrega de video, además de contar una estimación del indicador PDU. Por más información referente a los resultados obtenidos se invita al lector a referirse a [3].

Capítulo 5

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

5.1. Introducción

En este capítulo se describe la adaptación y aplicación de un modelo paramétrico de estimación de calidad de video, generado en el marco del proyecto VQI. Se tomó como base el modelo presentado en [30]. 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. 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 5.1 se presenta la formulación general del modelo,

$$MOSp = 1 + IcIp \quad (5.1)$$

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:

Capítulo 5. Aplicación de un modelo paramétrico de estimación de calidad de video en TV Digital

$$MOSp = 1 + Ic \quad (5.2)$$

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

5.1.1. Ic - Degradación asociada al proceso de codificación

En el modelo paramétrico seleccionado, Ic se calcula según la siguiente ecuación [30]:

$$Ic_{Pred} = v_3 \left(1 - \frac{1}{1 + \left(\frac{ab}{v_4} \right)^{v_5}} \right) \quad (5.3)$$

Donde,

$$v_3 = 4 + 4(f_{max} - f)(k_1s + k_2e^{-k_3(f_{max}-f)ab}) \quad (5.4)$$

$$v_4 = c_1s^{c_2} + c_3 \quad (5.5)$$

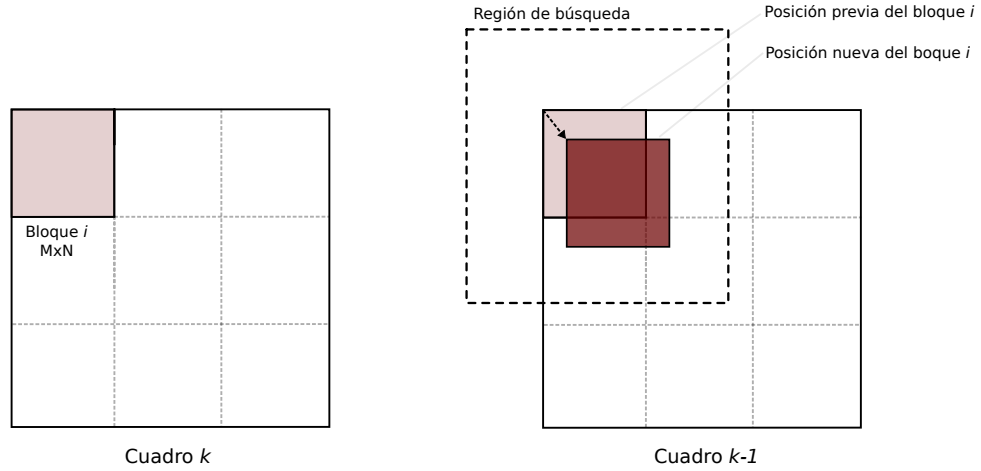
$$v_5 = c_4s^{c_5} + c_6 \quad (5.6)$$

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 al contenido del clip de video, denominada SAD (*Sum of Absolute Differences*). El algoritmo utilizado para su cálculo se ilustra en la figura 5.1, donde dada una secuencia de K imágenes se divide cada cuadro en I bloques de tamaño $M \times N$ (ej: 8×8) en forma de grilla. Luego se define una ventana de búsqueda y para cada bloque B_i de cada cuadro k se encuentra el bloque B_i^* que mejor lo represente dentro de esta ventana en el cuadro $k - 1$. La norma más frecuentemente utilizada para realizar la comparación entre bloques es el SAD (Sum of Absolute Differences), y se define según la siguiente expresión:

$$SAD(B, B^*) = \sum_{n=1}^N \sum_{m=1}^M |B(m, n) - B^*(m, n)| \quad (5.7)$$

Donde B y B^* son dos bloques de tamaño $M \times N$ y m y n son los índices horizontales y verticales, respectivamente. Siguiendo este procedimiento para cada uno de los bloques de cada cuadro de la secuencia se obtiene el parámetro s de acuerdo a la siguiente expresión:

$$s = \frac{1}{MN} \sum_{k=2}^K \sum_{i=1}^I SAD(B_{k,i}, B_{k-1,i}^*) \quad (5.8)$$

Figura 5.1: Ilustración del método de cálculo de la característica s (SAD)

A partir de este tipo de procedimientos es posible generar los vectores de movimiento, previamente mencionados en el capítulo 3. Éstos son utilizados en el área de compresión de video con el objetivo de disminuir la redundancia temporal de una secuencia de imágenes. Incluso se puede pensar s como el valor medio tras realizar el promedio de los valores de píxel 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 [85] muestra una alta correlación con los parámetros v_4 y v_5 , de allí es que surgen las ecuaciones 5.5, 5.6. Adicionalmente, tal y como se muestra en la tabla 4.1 se registró el promedio de amplitudes de los vectores de movimiento. Dicho valor actualmente no se utiliza para el cálculo de I_c aunque no se descarta para futuras versiones.

La idea intuitiva detrás de la ecuación 5.3 es que ésta modele la componente de calidad I_c 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 I_c . 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 I_c , respectivamente.

A continuación se describe el proceso de calibración de los coeficientes del modelo: c_1 , c_2 , c_3 , c_4 , c_5 , c_6 , k_1 , k_2 , k_3 y los resultados obtenidos en el proceso de entrenamiento del modelo.

Capítulo 5. Aplicación de un modelo paramétrico de estimación de calidad de video en TV Digital

Etiquetado de muestras

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. Luego, se procedió a aplicar la ecuación 5.9 para cada uno de los clips. Recordar que para estos videos I_p es igual al unidad.

$$I_c = MOS - 1, \quad (5.9)$$

siendo MOS el resultado de los ensayos subjetivos descritos en el capítulo 4 para cada clip. De esta manera se etiquetó cada uno de los clips con su respectivo I_c subjetivo.

Extracción de características

Las características identificadas como determinantes por el modelo de calidad utilizado (ecuación 5.3) son: bitrate, resolución y s (SAD). Para extraer el bitrate y la resolución de los clips codificados se utilizó la herramienta *FFmpeg*. Mientras que para extraer el SAD se utilizó algunas de las herramientas presentadas en [86], puntualmente el *plugin* denominado como *MVtools_v2* con *AVISynth*¹ y *VirtualDub*². Específicamente se utilizó la función “MAnalyse”, que por defecto utiliza tamaños de bloques de 8x8 para realizar el proceso de calculo del parámetro s y la estimación de los vectores de movimiento.

Calibración de coeficientes

El valor del coeficiente v_3 de la ecuación 5.3 se tomó constante con el valor de 4. Esto es dado que f corresponde al máximo frame rate utilizado (f_{max}), anulando así el segundo término de la ecuación 5.4. Al anular este segundo termino en esta etapa no fue necesario calibrar las constantes k_1, k_2, k_3 . Respecto del cálculo de a , en la tabla 5.1 se muestra el parámetro a extraído de [29] para distintas resoluciones.

Tabla 5.1: Valores de parámetro: a

Resolución	SD	VGA	CIF	QCIF
a	1	1,4	3,2	10,8

Con el fin de mantener la notación se mantuvo el valor de a para la resolución SD en la unidad (de ahora en más a_{SD}). Dado que el modelo inicial [29] no incluía la opción HD en la tabla 5.1 no figura esta resolución. Por lo tanto, fue necesario calibrar el valor de a para resolución HD, de ahora en más a_{HD} .

Con los valores de a_{SD} , b y s calculados para cada uno de los clips y sus correspondientes etiquetas (I_c) se procedió a calibrar v_4 , v_5 y a_{HD} ajustando la

¹En línea, http://avisynth.nl/index.php/Main_Page

²En línea, <http://www.virtualdub.org/>

5.1. Introducción

curva de la ecuación 5.3 minimizando el RMSE (*Root Mean Square Error*) para todos los videos del grupo de entrenamiento. El valor de a_{HD} obtenido fue de $a = 0,36957$ manteniendo la monotonía decreciente que se puede observar en la tabla 5.1 en función de la resolución.

Una vez obtenidos los valores de v_4 y v_5 se realizó el mismo procedimiento ajustando las curvas de las ecuaciones 5.5 y 5.6. En el proceso previo al ajuste, observando los datos se decidió tomar c_2 y c_5 con valor unitario, con el fin de ajustar una familia de rectas. En la tabla 5.2 se pueden observar los valores obtenidos para los coeficientes c_1 - c_6 .

Tabla 5.2: Parámetros del modelo paramétrico: c_1 - c_6

c_1	c_2	c_3	c_4	c_5	c_6
0,46	1	-0.48	0,16	1	0,51

Resultados Obtenidos

En la figura 5.2 se muestra la dispersión entre $I_{C_{Pred}}$ e I_C sobre los clips de video de entrenamiento seleccionados. Se obtuvo una correlación de Pearson (PLCC) de 0.8385 y un RMSE de 0.3356.

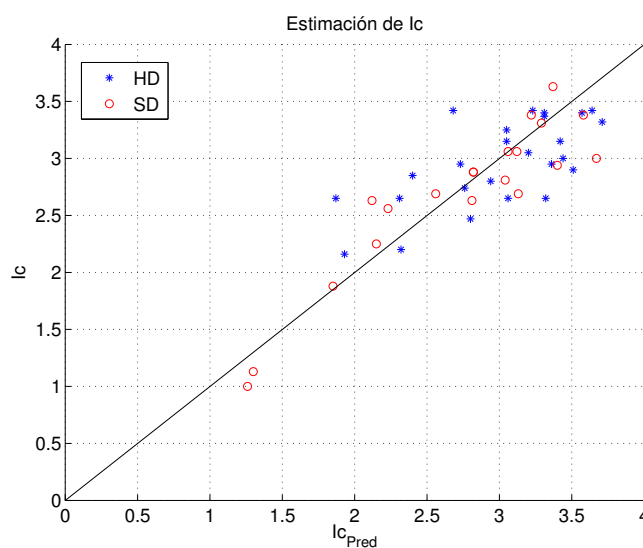


Figura 5.2: Estimación de I_C sobre la base de datos de entrenamiento

5.1.2. I_p - 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

Capítulo 5. Aplicación de un modelo paramétrico de estimación de calidad de video en TV Digital

I , P o B . Como ya ha sido señalado 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 [82] se propone un modelo que utiliza las ecuaciones 5.10 y 5.11. En esta tesis se denomina a este modelo de estimación de Ip por el nombre $Ip_{Inicial}$.

$$Ip_{Inicial} = \frac{1}{1 + kP_w} \quad (5.10)$$

$$P_w = x_1I + x_2P + B \quad (5.11)$$

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 .

Los valores brindados en [82] para los coeficientes k , x_1 , x_2 se presentan en la tabla 5.3.

Tabla 5.3: Valores para los coeficientes k , x_1 , x_2

k	x_1	x_2
21,5	5,7	26,9

Con el fin de medir el desempeño del modelo se calculó el valor subjetivo de Ip para aquellos clips de video del conjunto de entrenamiento *Set I HD* y *Set I SD* con pérdida de paquetes según la ecuación 5.12:

$$Ip = \frac{MOS - 1}{Ic} \quad (5.12)$$

Siendo MOS el resultado de los ensayos subjetivos descritos e Ic el calculado según la ecuación 5.9 para el mismo contenido con la misma codificación (*bitrate*) y sin pérdidas de paquetes.

A continuación se describen los resultados obtenidos. En la figura 5.3 (a) se puede observar Ip en función de P_w (ecuación 5.11). Mientras que en la figura 5.3 (b) se observa la dispersión entre Ip y su predicción $Ip_{Inicial}$ (ecuaciones 5.10 y 5.11). Para ello se utilizó todos los videos con $Ip < 1$ del conjunto de entrenamiento *Set I* con los valores mencionados en la tabla 5.3. Como resultado se obtuvo una PLCC de 0,837 y un RMSE de 0,141.

5.1. Introducción

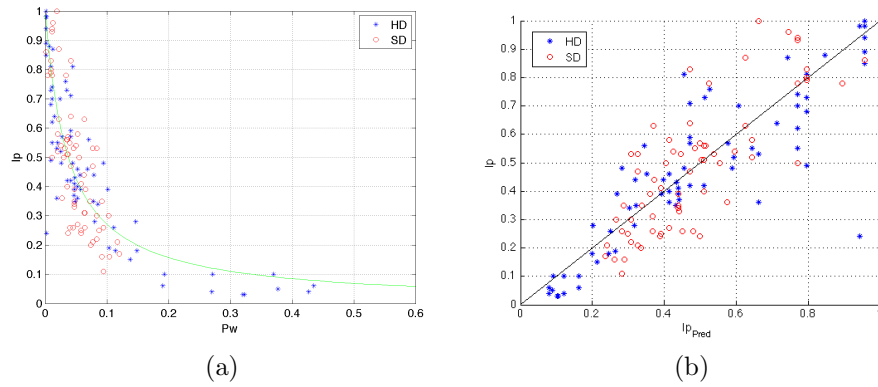


Figura 5.3: Ilustración de los resultados en la estimación de I_p al utilizar las ecuaciones 5.10 y 5.11 sobre los videos del conjunto de entrenamiento *Set I* con I_p menor a la unidad, en (a) se puede observar I_p en función de P_w y en (b) la dispersión entre I_p y su predicción $I_{p_{pred}}$

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 por que mantenerse.

Con este nuevo enfoque se propone calcular P_w de acuerdo a la ecuación 5.13:

$$P_w = \frac{C_A}{C_T} \quad (5.13)$$

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 I_p ($I_{p_{pred}}$) se evaluaron dos curvas de ajuste presentadas en las ecuaciones 5.14 y 5.15.

Capítulo 5. Aplicación de un modelo paramétrico de estimación de calidad de video en TV Digital

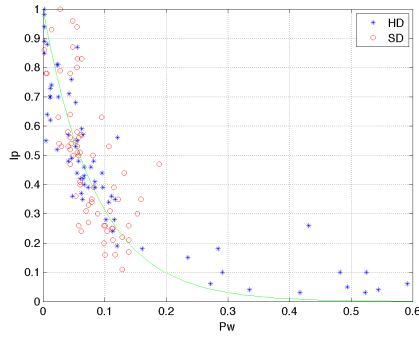
$$Ip_{Exp} = e^{-\alpha P_w} \quad (5.14)$$

$$Ip_{Pol} = \frac{1}{1 + \beta P_w} \quad (5.15)$$

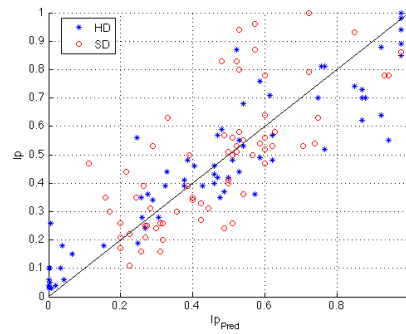
Donde α y β son los parámetros de calibración, respectivamente. En la figura 5.4 se muestran los resultados obtenidos tras aplicar ambas curvas de ajuste al nuevo modelo de P_w (ecuación 5.13). Al igual que en el caso anterior se utilizaron todos los videos con $Ip < 1$ de los conjuntos *Set I HD* y *Set I SD*. En la tabla 5.4 se muestran los coeficientes de calibración obtenidos en el ajuste.

Tabla 5.4: Valores para los coeficientes α , β

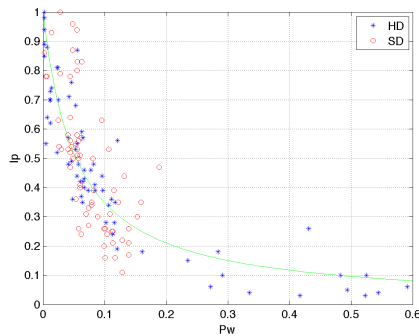
α	β
11,615	18,783



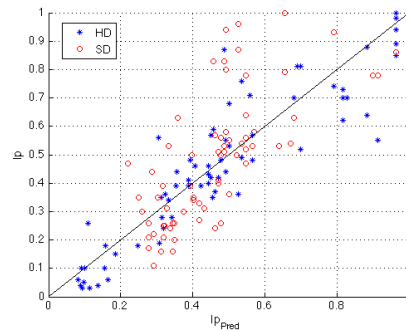
(a) Ajuste Exponencial



(b) Ip vs Ip_{Exp}



(c) Ajuste polinómico



(d) Ip vs Ip_{Pol}

Figura 5.4: Ilustración de los resultados en la estimación de Ip , en (a) se presenta el ajuste exponencial, en (b) se tiene la dispersión de Ip en función de su predicción Ip_{Exp} , en (c) se presenta el ajuste polinómico, en (d) se tiene la dispersión de Ip en función de su predicción Ip_{Pol}

5.1. Introducción

Tabla 5.5: Comparación de los resultados obtenidos para los modelos de estimación de I_p propuestos

Modelo	PLCC	RMSE
$I_{pInicial}$	0,837	0,141
I_{pExp}	0,844	0,146
I_{pPol}	0,832	0,143

En la tabla 5.5 se presentan la PLCC y el RMSE obtenido para cada uno de los modelos de estimación de I_p propuestos incluyendo el modelo Inicial. Se puede observar que los modelos de estimación de I_p utilizados tienen un desempeño similar sobre el conjunto de datos seleccionado. Sin embargo, cabe destacar que los modelos I_{pExp} e I_{pPol} extienden el uso del modelo Inicial a otros tipos de GOP , no solo al utilizado para la codificación de las secuencias seleccionadas.

5.1.3. MOSp - Predicción del MOS

Una vez obtenidos los coeficientes de los modelos asociados a I_c e I_p se puede obtener una estimación del MOS para una secuencia dada tras aplicar la ecuación general 5.1. En la tabla 5.6 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 [82], los valores de PLCC y RMSE presentados en esta tabla fueron extraídos de [82]. Por otro lado los modelos $MOSp_{Exp}$ y $MOSp_{Pol}$ utilizan el mismo modelo para I_c y distintos modelos de I_p . Por un lado el modelo $MOSp_{Exp}$ utiliza las ecuaciones 5.14 y 5.13 para realizar la estimación de I_p , mientras que el modelo $MOSp_{Pol}$ utiliza las ecuaciones 5.15 y 5.13. 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.

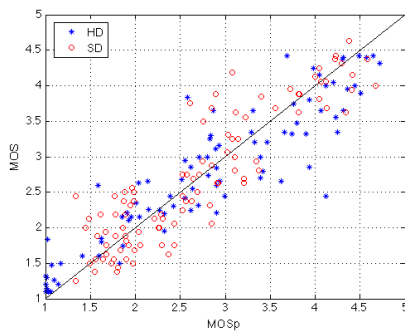
Cabe resaltar el hecho de que al aplicar el modelo de I_c en el *Set II* para la característica s asociada al SAD se obtuvieron valores menores a los esperados, fuera del rango probado. Particularmente estos valores se obtuvieron para secuencias con escenas de tipo informativo con muy baja actividad temporal. Este hecho repercutió en el cálculo del coeficiente v_4 , arrojando valores negativos. Esto se debe a que en primera instancia no se previó valores de s tan bajos en la ecuación 5.5 del modelo de I_c . Para mitigar este hecho y aplicar el modelo sobre estas secuencias se decidió saturar la variable v_4 con un umbral inferior, de valor 0,1. Este hecho se tendrá en cuenta para futuras versiones del modelo asociado a I_c y será tomado en consideración para futuras evaluaciones subjetivas.

En la tabla 5.6 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

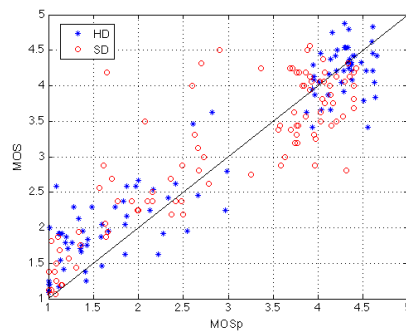
Capítulo 5. Aplicación de un modelo paramétrico de estimación de calidad de video en TV Digital

Tabla 5.6: Comparación de los resultados obtenidos en entrenamiento y validación de los modelos seleccionados

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

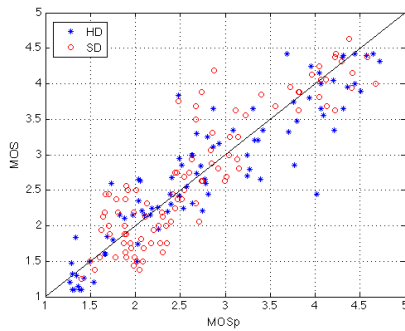


(a) Set I HD & Set I SD

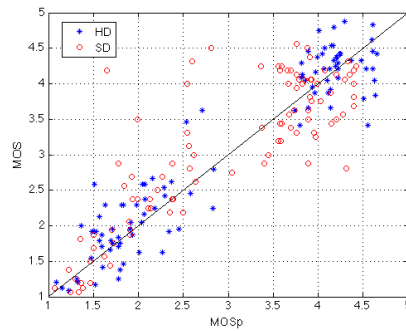


(b) Set II HD & Set II SD

Figura 5.5: Resultados obtenidos por el modelo $MOS_{p_{Exp}}$ en los conjuntos (a) Set I y (b) Set II, respectivamente.



(a) Set I HD & Set I SD



(b) Set II HD & Set II SD

Figura 5.6: Resultados obtenidos por el modelo $MOS_{p_{Pol}}$ en los conjuntos (a) Set I y (b) Set II, respectivamente.

de TV utilizada para registrar el contenido (*transport stream* recibido). Por otro lado es alentador el rendimiento de los modelos $MOS_{p_{Pol}}$ y $MOS_{p_{Exp}}$ en el Set II. En la figura 5.5 se presenta la dispersión entre el MOS y su predicción para el modelo $MOS_{p_{Exp}}$ aplicado en el conjunto Set I y luego en el conjunto Set II. Análogamente en la figura 5.6 se presenta la dispersión entre el MOS y su predicción para el modelo $MOS_{p_{Pol}}$.

5.1. Introducción

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.

Capítulo 6

Evaluación de calidad de video en un sistema de monitorización de la señal de TV Digital

6.1. Introducción

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.

En este capítulo se describe brevemente el proyecto de SMTVD. Se comienza con una descripción de la arquitectura geo-distribuida del sistema para luego desarrollar en detalle mi participación directa en el área de estimación de calidad. Particularmente, se describe el análisis, selección e integración de los modelos de evaluación de calidad de video utilizados por el sistema de monitorización. Además se evalúa una serie de indicadores que permiten medir el grado de severidad de distintos artefactos.

6.2. Flujo de trabajo de una estación de Televisión: *puntos de control de calidad*

En la cadena de trabajo de una estación de televisión se pueden distinguir, en principio, tres etapas donde la evaluación de calidad de video y audio es importante:

- 1- Dentro de los estudios de filmación, donde se genera la señal.
- 2- En la cadena de distribución, donde la señal puede atravesar distintos medios de comunicación así como distintos operadores de red.
- 3- En el sitio de recepción, donde degradaciones asociadas a la transmisión de la señal, por ejemplo interferencia en el canal de radiofrecuencia, provoquen errores que conlleven a la pérdida de paquetes en el flujo de transporte o *transport stream* recibido.

Dentro de los estudios de filmación es donde la calidad de la señal puede controlarse de mejor manera. Más allá de la presencia y la experiencia de los ingenieros en la estación existen herramientas comerciales automatizadas disponibles de fabricantes reconocidos. Luego la programación generada en los estudios es codificada y enviada para su posterior emisión a través de la cadena de distribución. Dicho proceso puede verse en la figura 6.1 donde se puede observar como la señal de video y audio se codifica por medio de un *encoder* o codificador. Una vez que la programación es codificada la calidad se degrada hasta cierto punto. Por otro lado la cadena de distribución es la red sobre la cual el programa previamente codificado llega a las instalaciones de transmisión. Usualmente se encuentra bajo el control del mismo radiodifusor o un operador de red. Si bien en ambos casos la cadena puede controlarse pueden existir diferencias entre la señal entrada y la de salida (en la figura 6.1 la señal codificada y la señal distribuida) cuando existen problemas de transmisión en la distribución.

Por último se encuentra la evaluación de calidad en el punto de recepción final, es decir, en los hogares de los televidentes dentro del área de cobertura de la estación de Televisión. Aún dada la importancia de la tarea de realizar un control de calidad en este punto, en los proyectos que ha llevado a cabo el VQEG en la última década ha sido complejo conseguir mediciones precisas. Una de las dificultades ha sido el hecho de que los modelos objetivos más precisos son los de referencia completa (FR), pero en los puntos de recepción, obviamente, sólo la señal recibida está disponible. Por lo tanto, no ha sido posible utilizar estos métodos para comparar la señal recibida con la original. En el marco del proyecto SMTVD, se ha propuesto un mecanismo que supera estas limitaciones, mediante una combinación de modelos con referencia completa y modelos sin referencia, según se explica en los siguientes párrafos.

En la figura 6.2 se ilustra el flujo de trabajo previamente descrito, desde la generación de las señales de video y audio hasta su visualización en los receptores de TVD en el área de cobertura de la estación de TV. Además en esta imagen se resalta qué tipo de degradaciones son introducidas en cada etapa y los puntos

6.2. Flujo de trabajo de una estación de Televisión: *puntos de control de calidad*

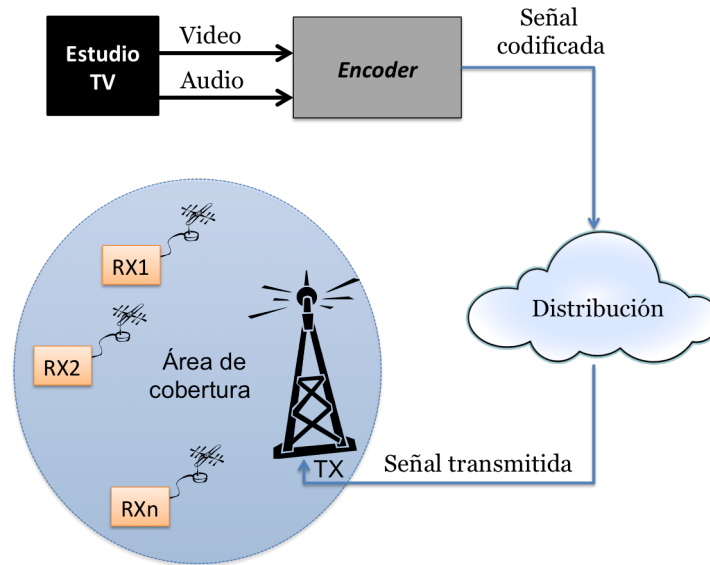


Figura 6.1: Etapas principales en la cadena de TVD donde es importante realizar un control de calidad

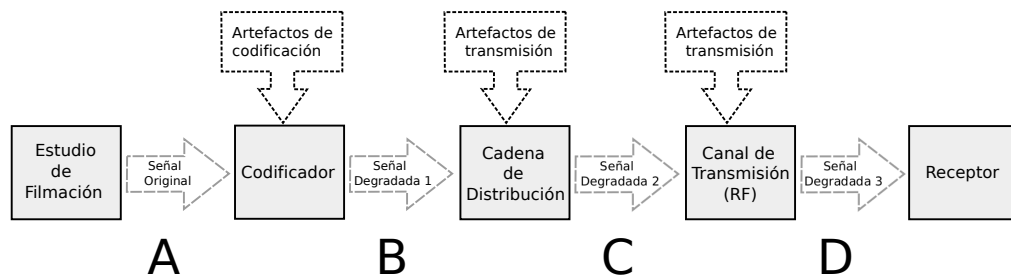


Figura 6.2: Puntos donde realizar una evaluación objetiva de la calidad de la programación

donde se puede realizar una evaluación de calidad, A, B, C y D respectivamente. Dado que las señales en los puntos A y B habitualmente se encuentran en el mismo lugar con el fin de obtener una estimación de la degradación de calidad percibida en el punto B (luego de la codificación) respecto de la señal original (Punto A) tiene sentido utilizar modelos FR. Como se ha mencionado existen modelos FR normalizados para tal tarea. Por otro lado modelos de tipo RR pueden ser utilizados para comparar la señal en el punto C al final de la cadena de distribución con la señal degradada en B. En este caso la cadena de distribución debe proveer un canal auxiliar para enviar información parcial de la señal. Por último, la señal recibida por los receptores de TV Digital es la señal del punto D. En este caso los únicos métodos aplicables son los NR dada la ausencia de posibilidad de comparar la señal con puntos anteriores. Cabe resaltar que estos métodos en general son menos precisos en términos de correlación con los resultados subjetivos de calidad. Es importante notar que si no hay pérdidas en la red de distribución y si la señal se recibe de manera satisfactoria en recepción, las señales digitales en los puntos

Capítulo 6. Evaluación de calidad de video en un sistema de monitorización de la señal de TV Digital

de recepción $Rx_1, Rx_2, \dots, Rx_n$ en la figura 6.1 serán idénticas, sin degradaciones debidas al proceso de transmisión (*error-free*). Con respecto a la figura 6.2 esto aplica a las señales de los puntos D y C, en ausencia de errores en distribución y transmisión serán idénticas, donde el proceso de adquisición y codificación son los componentes que causan la degradación de calidad en las señales de video y audio.

Cabe destacar que cada canal de radiofrecuencia desde la antena transmisora a los receptores $Rx_1, Rx_2, \dots, Rx_n$ es diferente, puede tener distinta respuesta en frecuencia, patrones de pérdida por interferencia o efecto multicaminos (comúnmente denominado como *multipath*). De todas maneras puede ocurrir que este fenómeno no implique una diferencia en el programa recibido de TVD. Para ello, es común utilizar el *Bit Error Rate* (BER) presente en cada punto de recepción. Dependiendo de este valor el programa puede: (i) ser idéntico al programa transmitido si el BER se encuentra por debajo de cierto umbral, (ii) visualizarse completamente congelado si el BER se encuentra por encima del umbral y (iii) por último debido a ocasionales pérdidas de paquetes si el BER oscila alrededor del umbral pueden visualizarse distorsiones en la imagen como *slicing* o pérdida de macro bloques, además de afectación en el audio como saltos o silencios.

Por ello, con el objetivo de estimar la calidad percibida de video por los televidentes de la señal de TV Digital emitida por una estación de TV en su área de cobertura es necesario contar con un sistema geográficamente distribuido que optimice la utilización de la información disponible en los puntos de medida descritos anteriormente. Por otro lado al diseñar este sistema sería deseable minimizar el uso de modelos NR, tratando de utilizar modelos RR o FR comparando la señal recibida en los puntos de recepción con aquella lo más cercana posible a la señal original.

6.3. Sistema de Monitorización de la señal de TV Digital

El sistema ha sido diseñado para servir no solo 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 (Punto A en figura 6.2). En la figura 6.3 se puede observar una ilustración a alto nivel de la arquitectura del sistema propuesto.

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*) y de cada una de las señales decodificadas (por ejemplo, afectación de cuadros o slices I, P o B en cada uno

6.3. Sistema de Monitorización de la señal de TV Digital

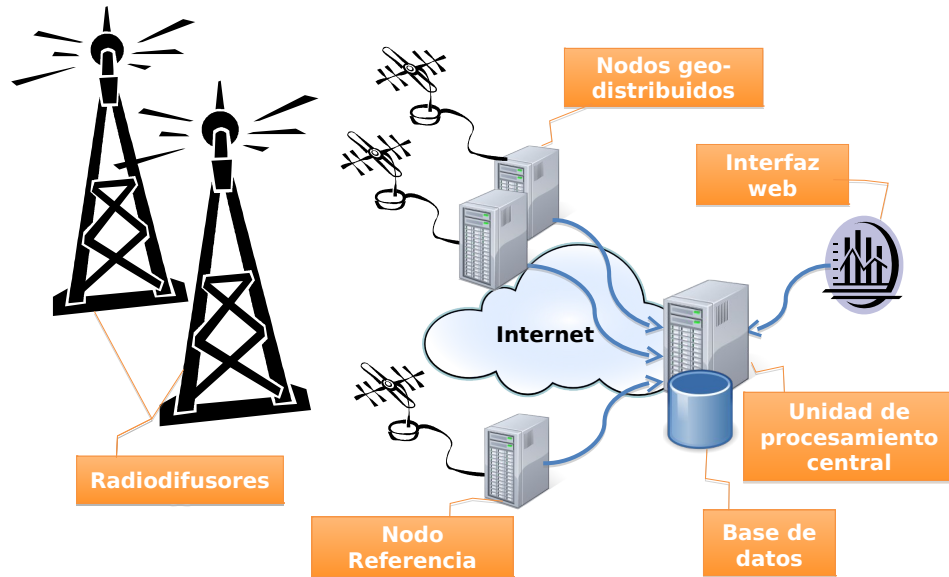


Figura 6.3: Ilustración a alto nivel de la arquitectura del sistema de monitorización en tiempo real de la señal de TVD

de los *elementary streams* recibidos). Por último, se realizan grabaciones de 10 segundos, donde se registra el *transport stream* 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 *transport streams* 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.

Capítulo 6. Evaluación de calidad de video en un sistema de monitorización de la señal de TV Digital

Sobre esta hipótesis de trabajo y en base a los puntos de control identificados anteriormente, la calidad obtenida para la programación en el nodo de referencia (punto D de la figura 6.2) es la misma que en el punto B. En tanto, con el objetivo de obtener una estimación de la calidad de video en el nodo referencia se pueden utilizar únicamente modelos NR, dado que no se tiene acceso a la señal original (punto A). Incluso pueden medirse distintos indicadores de calidad de servicio en la señal grabada, por ejemplo: *bitrate*, *framerate*, borrosidad, efecto de bloques, entre otros. Periódicamente, el nodo de referencia envía a la unidad de procesamiento central las distintas medidas y los estimadores de calidad.

Los nodos se comunican con la unidad de procesamiento central a través de Internet utilizando mecanismos de comunicación seguros. Periódicamente o por demanda, los nodos remotos envían a la unidad de procesamiento central las distintas medidas y el porcentaje de pérdida de paquetes detectado. 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) [87] 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. Luego, un modelo FR de estimación de calidad de video puede ser utilizado para obtener una estimación de la degradación de la calidad percibida en el nodo remoto. En la siguiente sección se especifican los modelos FR utilizados. Con el objetivo de obtener una estimación global de la calidad percibida de video en cada uno de los nodos remotos, el sistema central combina la estimación de calidad en el nodo de referencia (por medio de un modelo NR) y la estimación de la degradación percibida en los nodos remotos (por medio de modelos FR). Posteriormente todas estas medidas y cálculos son guardados en la base de datos y presentados en línea vía una portal web especialmente diseñado para esta tarea. Los archivos guardados en los nodos remotos pueden ser solicitados en cualquier momento y pueden ser visualizados utilizando una interfaz gráfica a modo de evidencia.

6.4. Modelo global de estimación de calidad de video

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 5.1, utilizada por el modelo paramétrico descrito en el capítulo 5. 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 o bien 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

6.4. Modelo global de estimación de calidad de video

referencia y la señal recibida en el nodo remoto (por medio de un modelo FR). En tanto, para el cálculo de I_p tiene sentido utilizar un modelo FR de estimación de calidad en vez de uno sin referencia.

A continuación se describe el cálculo de las componentes I_c e I_p en el sistema.

6.4.1. Cálculo de I_c

Con el fin de realizar el cómputo de I_c en el sistema SMTVD se utiliza el modelo paramétrico de la ecuación 5.3 presentado en el capítulo 5. Este modelo se aplica a los *elementary streams* de video asociados al *transport stream* 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. El modelo tiene como insumo el bitrate, resolución y SAD de cada una de las programaciones. El bitrate y la resolución particularmente se extraen del bitstream utilizando la aplicación *FFmpeg*¹. El SAD se extrae utilizando la aplicación *Virtualdub*² en conjunto con *FFmpeg*. Estas aplicaciones son invocadas de manera automática y corren sobre un Sistema Operativo Linux en el nodo referencia. Con este procedimiento se obtiene el I_c 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 5.1.

Análisis de tiempos de ejecución

En la tabla 6.1 se muestra el costo en términos de tiempo de ejecución de la implementación del modelo de estimación de I_c seleccionado y analizado en este trabajo. Cabe mencionar que la integración e implementación del modelo seleccionado de estimación de I_c 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.

Tabla 6.1: Tiempo de ejecución del modelo asociado a I_c

Indicador	HD	SD
I_c	13.3 seg	8.1 seg

Como puede verse, utilizando el hardware mencionado, I_c puede calcularse de manera continua y en tiempo real para señales en resolución SD. Es decir, realizando una estimación de I_c para cada clip registrado de 10 segundos, seis por minuto. Para señales HD, se podría tener un valor de I_c cada 15 segundos, cuatro veces por minuto, realizando solo la estimación sobre 10 de los 15 segundos.

¹*FFmpeg*, en línea, <https://www.ffmpeg.org>

²*Virtualdub*, en línea, <http://www.virtualdub.org>

Capítulo 6. Evaluación de calidad de video en un sistema de monitorización de la señal de TV Digital

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.

6.4.2. Cálculo de Ip

En el capítulo 5 se ha descrito el modelo paramétrico inicial empleado para la predicción de Ip , propuesto en [82]. Luego se extendió dicho modelo, obteniendo una mejora en desempeño en la estimación del MOS sobre grabaciones reales de TVD. En esta sección se muestra como al utilizar modelos FR se mejora el desempeño de los modelos paramétricos de estimación de Ip utilizados.

Acorde a las ecuaciones 5.13, 5.14 y 5.15 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 6.1:

$$Pw = 1 - \frac{1}{N} \sum_{i=1}^N Q(i) \quad (6.1)$$

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 embargo 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 con referencia completa (FR) 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 de estimación de calidad de video (VQA – *Video Quality Assessment*) así como de imágenes (IQA – *Image Quality Assessment*) que pueden ser utilizados para esta tarea [18]. En este trabajo, en el marco del proyecto SMTVD, se evaluaron los siguientes modelos, previamente descritos en el capítulo 2.

- **PSNR**: Peak Signal-to-Noise Ratio,
- **SSIM**: Structural Similarity Index,
- **MS-SSIM**: Multi-Scale Structural Similarity Index,

6.4. Modelo global de estimación de calidad de video

- **VIFp**: Visual Information Fidelity, pixel domain version,
- **PSNR-HVS**: Peak Signal-to-Noise Ratio tomando en cuenta la función de sensibilidad de contraste (CSF).
- **VQM *General***: Video Quality Metric, General model
- **VQM *LowBw***: Video Quality Metric, Low Bandwidth Model

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 6.2.

$$P_w = 1 - \frac{1}{N} \sum_{j=1}^N Q_{v_x}(i) \quad (6.2)$$

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 6.1 con 6.2 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 6.4 se observa el resultado de los modelos *SSIM*, *MS-SSIM* y *VIFp* para cada uno de los cuadros sobre 4 clips de videos 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* [42] propuesto por la NTIA y el modelo VQM *LowBw* [47]. 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>).

Capítulo 6. Evaluación de calidad de video en un sistema de monitorización de la señal de TV Digital

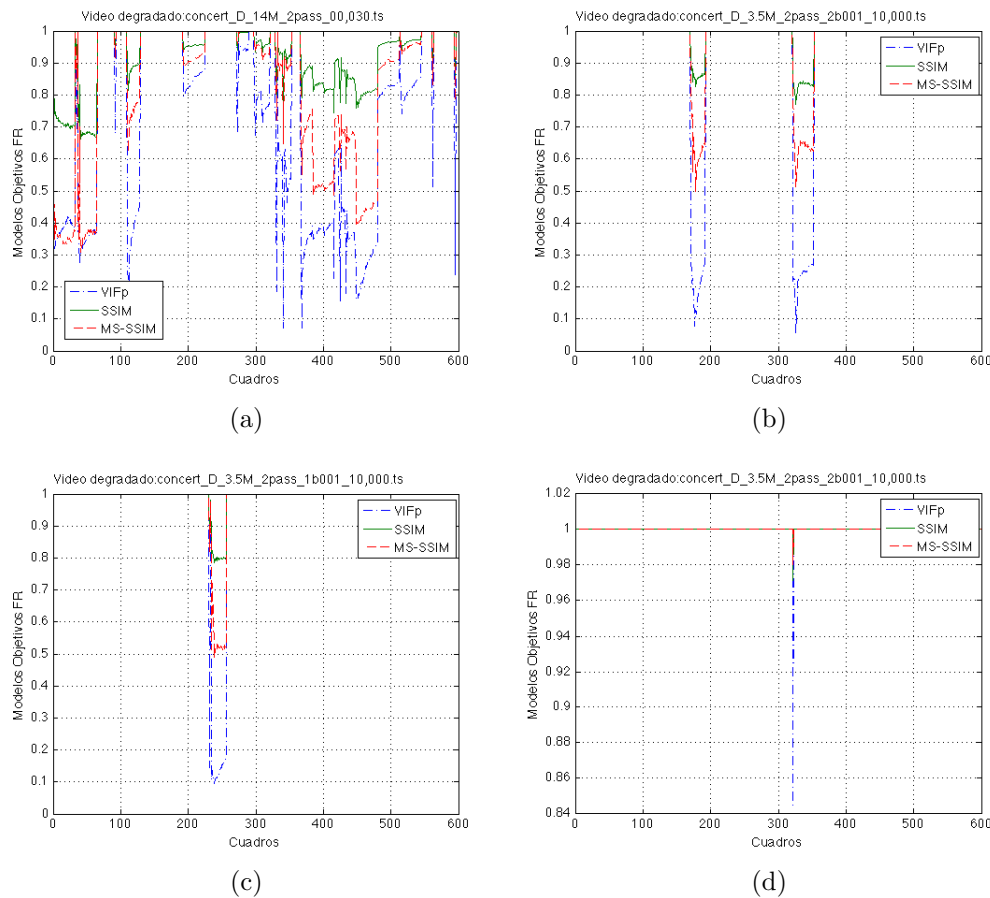


Figura 6.4: Resultado de los modelos VIFp, SSIM y MS-SSIM sobre cada uno de los 600 cuadros que conforman un clip de video de la base *Set I HD* con (a) distribución uniforme de pérdida de paquetes (b) dos ráfagas (*bursts*) de errores (c) una ráfaga de errores. Por último en (d) se observa un ejemplo donde solo se vio afectado un cuadro, en este caso se afectó un *slice* de tipo B

Aunque hoy en día no sea una medida aceptada de calidad percibida por observadores humanos [31] igualmente se decidió evaluar el PSNR por ser una medida históricamente utilizada. Recordando la formulación del PSNR para su cálculo entre dos imágenes, ecuación 2.2, 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 [88] para transmisiones

6.4. Modelo global de estimación de calidad de video

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 2.2 para cada par de cuadros y luego realizar un promedio entre la cantidad total de medidas obtenidas, se calcula el MSE, según la ecuación 2.1 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 2.2. Con esta aproximación se logra que el PSNR sea ∞ solo cuando los videos son idénticos, en cuyo caso $I_p = 1$. Por último, en adición al PSNR se 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.

Desempeño de los modelos propuestos para la estimación de I_p

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 I_p 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 ($I_p < 1$) 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 modelo objetivos se utiliza un ajuste no lineal [89] para convertir el resultado de los modelos a una escala lineal específica respecto del resultado subjetivo (I_p).

En este trabajo se utilizó la transformación no lineal presentada en 6.3:

$$I_{p_{pred}} = t_1 \left(0,5 - \frac{1}{1 + e^{(t_2(x-t_3))}} \right) + t_4 \quad (6.3)$$

Donde $I_{p_{pred}}$ es la predicción de I_p 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. Para calcular estos coeficientes se ajustó la curva de la ecuación 6.3 a los valores de I_p para cada uno de los clips seleccionados minimizando la raíz del error cuadrático medio (RMSE).

En la figura 6.2 se muestra la relación entre el valor de I_p y el resultado directo de los modelos objetivos utilizados (x) para cada uno de los clips de video empleados (izquierda), además se muestra la dispersión entre el valor de I_p y su

Capítulo 6. Evaluación de calidad de video en un sistema de monitorización de la señal de TV Digital

predicción tras aplicar la ecuación 6.3 (derecha). Con el símbolo “*o*” se identifican los videos en resolución SD y con “***” los videos en resolución HD.

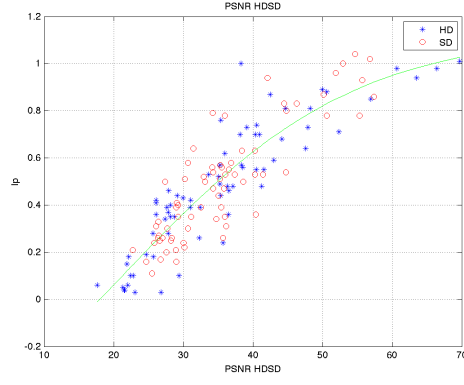
Idealmente en esta última si las predicciones fueran perfectas, los puntos deberían ubicarse sobre una recta a 45 grados. Sin embargo, se observa cierta dispersión, la que es medida para cada modelo y utilizada para realizar una comparación de desempeño. Con este fin se utilizó la correlación de Pearson y el RMSE entre los valores de I_{pPred} e I_p y las correlaciones de Spearman y Kendall entre el resultado x asociado a cada modelo y los valores de I_p . 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 la tabla 6.2 se muestra el desempeño de los modelos propuestos.

Tabla 6.2: Desempeño de los modelos FR/RR propuestos sobre los videos con pérdida de paquetes ($I_p < 1$) 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)

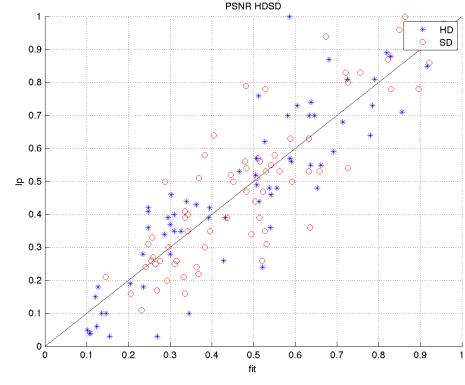
HD & SD	PLCC	SROCC	KROCC	RMSE	Tiempo (HD)	Tiempo (SD)
I_{pPSNR}	0.88825	0.87154	0.69602	0.11786	18.1 seg	4.3 seg
$I_{pVQMGeneral}$	0.82847	-0.8392	-0.6679	0.14369	6.9 min	78.7 seg
$I_{pVQMLowbw}$	0.90063	-0.89598	-0.73846	0.1115	8.9 min	103.7 seg
$I_{pMS-SSIM}$	0.89329	0.90466	0.74539	0.11533	40.1 seg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

En comparación con los modelos paramétricos NR inicialmente propuestos, cuyo desempeño se muestra en la tabla 5.5, se puede observar que los modelos I_{pPSNR} , $I_{pVQMLowbw}$, $I_{pMS-SSIM}$, I_{pSSIM} e I_{pVIFp} 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 [46] utilizando la Transformación z de Fisher, descrita en la sección 2.5. Para ello se convierte la PLCC a una variable z con distribución normal de acuerdo con la ecuación 2.17. El test estadístico asume la hipótesis nula (H_0) de que no hay diferencia significativa entre los coeficientes de correlación de Pearson (PLCC). A continuación se calcula la variable Z_{jk} según la ecuación 2.18. Donde z_j es la Transformación z de Fisher del modelo j y z_k del modelo k y N es la cantidad de muestras que coincide la cantidad de clips de video utilizada para calcular la PLCC. En este caso $N > 30$ por lo tanto se puede asumir una distribución normal, en consecuencia Z_{jk} determina el nivel de confianza (por ejemplo, si $Z_{jk} > 1.96$ implica que el modelo j es estadísticamente mejor que el modelo k con un nivel de confianza del 95 %). En este caso los modelos $I_{pVQMLowbw}$, I_{pSSIM} e I_{pVIFp} muestran una mejora estadística respecto al modelo $I_{pInicial}$ con una nivel del confianza del 97 %. Este tipo de resultado concuerda con

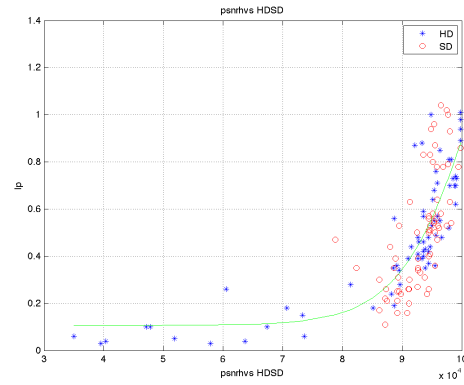
6.4. Modelo global de estimación de calidad de video



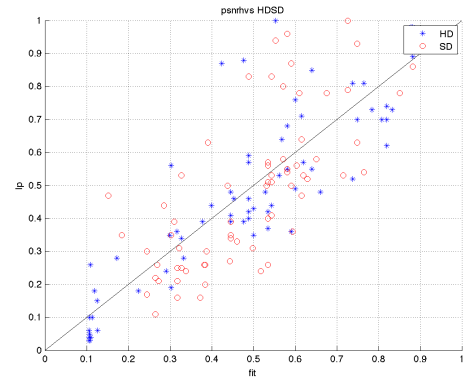
(a) $PSNR$



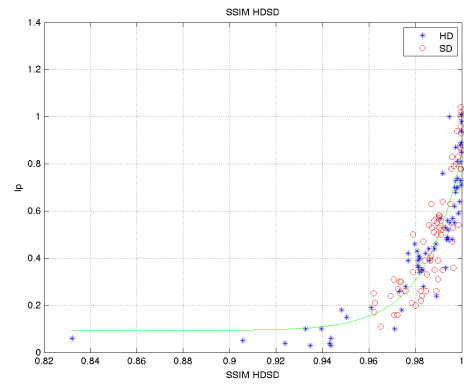
(b) Ip_{PSNR}



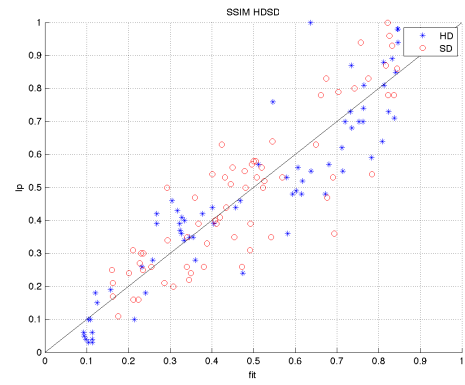
(c) $PSNR - HVS$



(d) $Ip_{PSNR-HVS}$

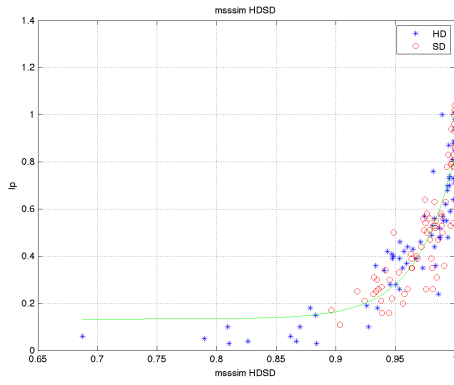


(e) $SSIM$

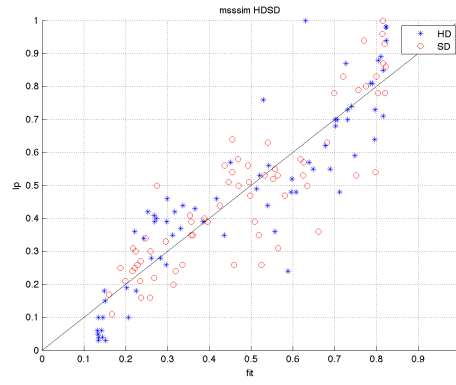


(f) Ip_{SSIM}

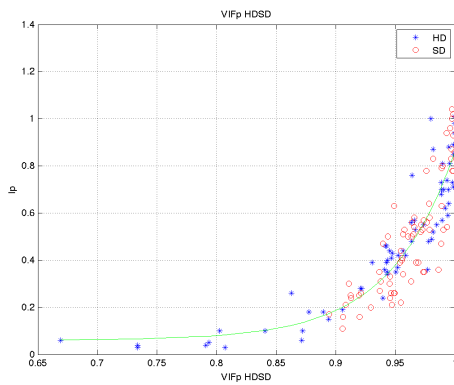
Capítulo 6. Evaluación de calidad de video en un sistema de monitorización de la señal de TV Digital



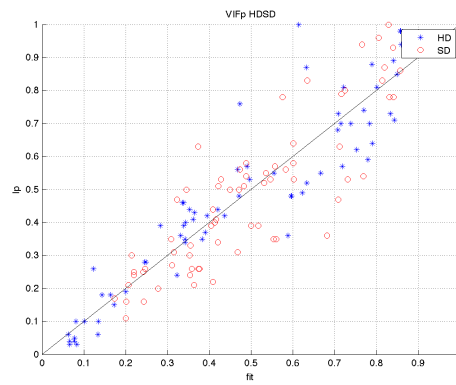
(g) $MS - SSIM$



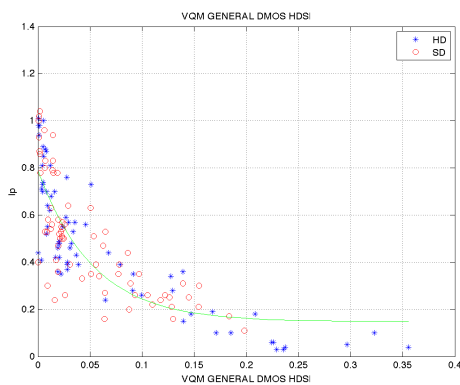
(h) $Ip_{MS-SSIM}$



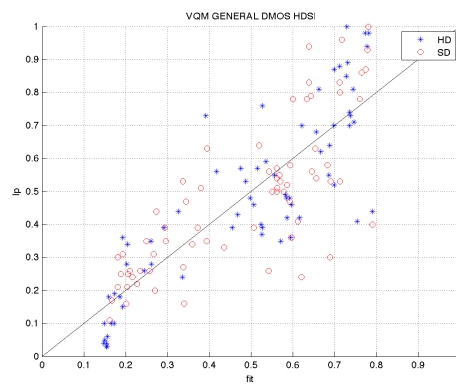
(i) $VIFp$



(j) Ip_{VIFp}



(k) $VQMGeneral$



(l) $Ip_{VQMGeneral}$

6.5. MOSp - Predicción del MOS

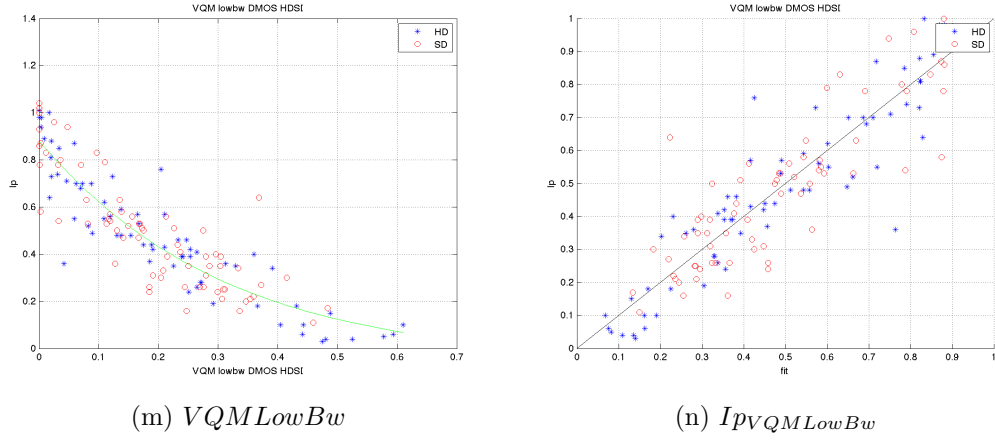


Figura 6.2: Desempeño de los modelos PSNR, PSNR-HVS, SSIM, MS-SSIM, VIFp, VQM “General” y VQM “LowBw”

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 I_p 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 I_{pSSIM} 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 I_p 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.

6.5. MOSp - Predicción del MOS

Con el fin de realizar la estimación del MOS empleando los distintos modelos propuestos se utiliza la formula global 5.1, con el componente I_c dado por la ecuación 5.9 y el componente I_p correspondiente a los distintos modelos de estimación de I_p . 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 I_p . 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 I_p .

Capítulo 6. Evaluación de calidad de video en un sistema de monitorización de la señal de TV Digital

En la tabla 6.3 se muestra el desempeño en términos de PLCC y RMSE obtenido en cada caso.

Tabla 6.3: Desempeño de los modelos FR/RR/NR propuestos sobre los clips de video con pérdida de paquetes TS ($I_p < 1$) 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
$MOSp_{VQMLowbw}$	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

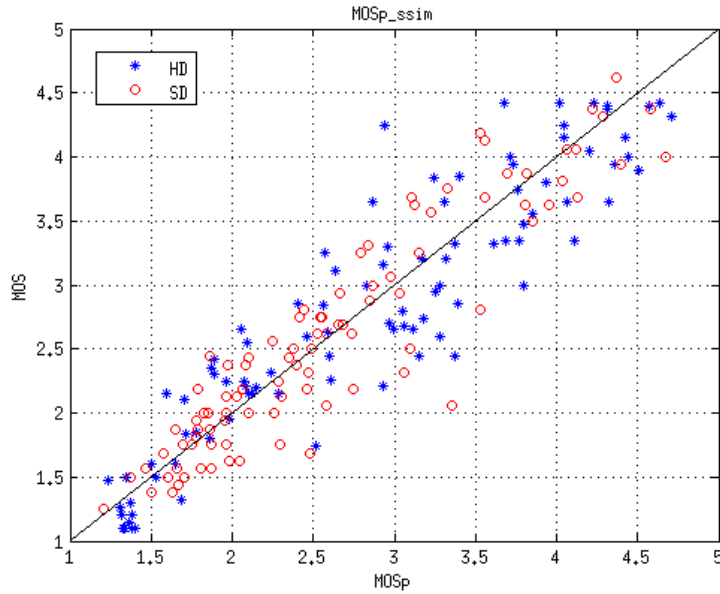


Figura 6.3: Dispersión MOS vs $MOSp_{SSIM}$ sobre el conjunto *Set I HD* y *Set I SD*

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

6.6. Alineación temporal de clips de video

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 6.3 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 5.6 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 %.

Hasta este punto se ha mostrado que al utilizar modelos FR/RR en la estimación de I_p se obtiene un mejor desempeño respecto de los modelos paramétricos NR propuestos. Esta comparación se realizó sobre los clips de video de los conjuntos *Set I HD* y *Set I SD*. Cabe destacar que los modelos FR/RR no se pueden aplicar sobre el conjunto independiente de clips de video *Set II HD* y *Set II SD* dado que estos videos no cuentan con una referencia. Aún así sí se pueden aplicar en el sistema de monitoreo. Como trabajo a futuro con el fin de extender la verificación y validación sobre secuencias reales 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.

6.6. Alineación temporal de clips de video

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 ambas secuencias. El algoritmo de alineación que se propone realiza la alineación temporal en 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.

Capítulo 6. Evaluación de calidad de video en un sistema de monitorización de la señal de TV Digital

Flujos de transporte (TS, *Transport Stream*)

Cada uno de los nodos graba el *Transport Stream* recibido en un momento establecido por el sistema. Cada TS contiene cada uno de los servicios que emite cada canal, por ejemplo un servicio HD, su espejo en SD, otro servicio SD y el servicio *one-seg* para receptores móviles. En lo que sigue se muestra el desglose de un TS grabado de Televisión Nacional Uruguay (TNU).

Duration: 00:00:09.50, **start:** 5702.557567, **bitrate:** 17323kb/s **Program 32960**

Stream # 0: 0 [0×201]: **Video:** h264 (High) ([27][0][0][0]/0×001B), yuv420p(tv, bt709), **1920x1080** (SAR 1:1 DAR 16:9), 25 fps, 25 tbr, 90k tbn, 50 tbc

Stream # 0:1 [0 × 202]: **Audio:** aac latm ([17][0][0][0]/0 × 0011), 48000 Hz, stereo, fltp

Program 32961

Stream # 0:2 [0×301]: **Video:** h264 (Main) ([27][0][0][0]/0×001B), yuv420p(tv, bt470bg), **720x576** (SAR 12:11 DAR 15:11), 25 fps, 50 tbr, 90k tbn, 50 tbc

Stream # 0:4 [0x302]: **Audio:** aac latm ([17][0][0][0]/0 × 0011), 48000 Hz, stereo, fltp

Program 32962

Stream # 0:3 [0×401]: **Video:** h264 (Main) ([27][0][0][0]/0×001B), yuv420p(tv, bt470bg), **720x576** (SAR 12:11 DAR 15:11), 25 fps, 50 tbr, 90k tbn, 50 tbc

Stream # 0:5 [0 × 402]: **Audio:** aac latm ([17][0][0][0]/0 × 0011), 48000 Hz, stereo, fltp

Program 32984

Stream # 0:6 [0×101]: **Video:** h264 (Constrained Baseline) ([27][0][0][0]/0×001B), yuv420p, 320x180 (SAR 1:1 DAR 16:9), 14.99 fps, 29.97 tbr, 90k tbn, 29.97 tbc

Stream # 0:7 [0 × 102]: **Audio:** aac latm ([17][0][0][0]/0 × 0011), 48000 Hz, stereo, fltp

EL TS fue grabado en dos ubicaciones distintas. Uno de los flujos fue captado por el nodo de referencia mientras el segundo en un nodo remoto del sistema de monitorización. Se puede observar que el TS contiene 4 programas o servicios, el primer programa contiene una señal en HD (1920 × 1080) con compresión h.264/AVC asociada al PID 0 × 201 y por otro lado una señal de audio con el PID 0 × 202. Además se puede ver que se tienen dos servicios en SD y uno *one-seg* para dispositivos móviles.

6.6. Alineación temporal de clips de video

Extracción de los flujos elementales (ES, Elementary Streams)

Con el objetivo de obtener cada uno de los flujos elementales (ES) y obtener las secuencias individuales de video se utiliza la herramienta *FFmpeg*. En primer lugar se extrae el *Elementary Stream* que contiene el clip de video que se desea procesar utilizando el siguiente comando:

```
ffmpeg -<referencia>.ts -map 0:i:<pid> -codec:v copy -f mpegts -  
y <referencia>_<pid>.es
```

Donde, <referencia> es el nombre del TS grabado en el nodo de referencia y <pid> es el PID del servicio que se desea extraer, por ejemplo 0×201 . Este comando se aplica para el TS de referencia y para el TS degradado. En lo que sigue se utilizan solo los *Elementary Streams* generados.

Generación de secuencias de cuadros en formato YUV

Una vez que se obtienen los *ES* se procede a generar las secuencias descomprimidas en formato YUV. Específicamente se utiliza el formato de pixel YUV420. Dado que alguno de los modelos FR propuestos necesitan que el largo y el ancho de las cuadros sea múltiplo de 16 pixeles en el proceso decodificación de ser necesario se agregan bandas negras. Por ejemplo una resolución de 1920×1080 se debe llevar a una de 1920×1088 , en este caso se introducen bandas negras arriba y abajo con un ancho de 4 pixeles. Por otro lado se debe forzar a que se mantenga la tasa de cuadros a la salida, en caso de que se pierda la información de un cuadro completo se debe repetir el cuadro disponible anterior. Para efectuar estas operaciones se utiliza el siguiente comando:

```
ffmpeg -i <elementary_stream>.es -vcodec rawvideo -pix_fmt  
yuv420p -vf pad=<new_WIDTH>:<new_HEIGHT>:0:0:black -vcodec  
rawvideo -vsync 1 -an -y -f framemd5 <elementary_stream>.md5  
- y <elementary_stream>.yuv
```

Donde, <new_WIDTH> y <new_HEIGHT> son el nuevo ancho y largo, respectivamente y <elementary_stream>.yuv es el YUV extraído. Por último se genera un archivo (<elementary_stream>.md5) 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 secuencia de video descomprimida en formato YUV será utilizada por los modelos objetivos FR para efectuar la estimación de calidad mientras el segundo archivo, que contiene el identificador (CRC) de cada uno de los cuadros decodificados, será utilizado para realizar el proceso de alineación de cuadros, que se describe a continuación.

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

Capítulo 6. Evaluación de calidad de video en un sistema de monitorización de la señal de TV Digital

correctamente el cuadro completo esta perdida (por ejemplo, Vectores de Movimiento y/o Macro Bloques perdidos). En este escenario los CRC de los cuadros afectados de la secuencia degradada serán distintos a 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 6.4 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.

Tabla 6.4: 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	fd6fef78c2a49418acb4089efddbd0d

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 hacia 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 los modelos FR, para ello se utiliza el software VQMT. Éste se debió adaptar para además de ingresar la secuencia original y degradada en formato descomprimido (YUV) también incluir estos *offsets*. Internamente en el código se modificó el puntero incorporando el *offset* en bytes correspondiente para cada una de estas secuencias. Con esto se calibran internamente los cuadros y se aplican los distintos modelos con los cuadros perfectamente alineados. Para ello se debe pasar el número de bytes correspondiente a cada *offset*. En este caso cada uno de los cuadros ocupa en memoria 622080 bytes (720×576 , formato de pixel: YUV420), este valor se extrae del archivo `<elementary_stream>.md5`. Por lo tanto los *offsets* en este ejemplo serán de 2×622080 bytes y 622080 bytes, respectivamente. De acuerdo al ejemplo en este caso el software comparará el cuadro número 2 de la secuencia original con el cuadro número 1 de la secuencia degradada, identificados con †, hasta comparar el cuadro número 12 de la secuencia original con el cuadro número 11 de la secuencia degradada identificados con ★. En tanto para estas secuencias se procesarán 11 cuadros.

6.7. Indicadores MOAVI

MOAVI³ (*Monitoring Of Audiovisual Quality by Key Indicators*) es un grupo de colaboración abierto para el desarrollo de modelos NR para monitorización de calidad audiovisual promovido por el VQEG. El objetivo de este grupo es desarrollar un conjunto de *Key Indicators* o indicadores clave que describan la calidad del servicio en general y de este modo poder seleccionar un conjunto reducido dependiendo de la aplicación objetivo. Estos indicadores permiten medir de manera objetiva la magnitud de distintos artefactos asociados al proceso de compresión y adquisición para cada cuadro de un clip de video. Luego pueden ser consolidados para obtener una estimación del nivel perceptible de cada artefacto, promediando entre la secuencia de cuadros. Como complemento de las medidas de calidad percibida de video en el sistema de monitorización se incluyeron los indicadores MOAVI descritos en la tabla 6.5. Estos se medirán en el nodo de referencia y un ejemplo de su cálculo se puede ver en <http://vq.kt.agh.edu.pl/metrics.html>.

³En línea, <http://www.its.bldrdoc.gov/vqeg/projects/moavi/moavi.aspx>

Capítulo 6. Evaluación de calidad de video en un sistema de monitorización de la señal de TV Digital

Estos indicadores pueden ofrecer información valiosa de la señal de video recibida. El sistema SMTVD puede utilizar esta información para disparar alarmas cuando alguno de estos indicadores se encuentra fuera de cierto rango. Por otro lado el registro sistemático de estos indicadores puede proveer información estadística de características específicas de las señales monitorizadas (por ejemplo, porcentaje de tiempo de pantalla en negro o con *pillarboxing*).

Sin embargo en este trabajo se intentó evaluar su uso con el fin de lograr una mejora en la estimación de I_c respecto del modelo paramétrico utilizado en la ecuación 5.3. Particularmente estas técnicas se utilizaron en la tesis doctoral de Piotr Romaniak [90] con el fin de probar la siguiente hipótesis:

“ Es posible evaluar la calidad perceptual en tiempo real del contenido de un video afectado por artefactos asociados a la adquisición y reducción de bitrate por medio de indicadores de tipo sin referencia.”

En dicho trabajo se consiguió probar esta hipótesis al proponer un conjunto de indicadores sin referencia y medir su desempeño en términos de correlación con la opinión subjetiva. La verificación de los resultados se llevó a cabo mediante ensayos subjetivos sobre un conjunto de clips de video con diversidad de contenido. Específicamente a partir de los resultados obtenidos sobre videos con compresión H.264 con reducción de *bitrate* y definición estándar (SD) se concluyó que estos indicadores pueden ser utilizados para realizar una evaluación confiable de calidad de video en tiempo real.

Con el objetivo de verificar que los tiempos de ejecución de los indicadores MOAVI permitan un procesamiento sistemático en tiempo real por el sistema SMTVD se realizó una evaluación de todos los indicadores previamente descritos. Para ello se utilizó un hardware con un procesador Intel Core i7-4702HQ con 16GB RAM y sistema operativo Ubuntu (Linux). Al igual que para el cómputo de I_c por medio del modelo paramétrico de la ecuación 5.3 se utilizaron grabaciones de 10 segundos registradas por el sistema SMTVD en resoluciones HD y SD. Los resultados obtenidos se presentan en la tabla 6.6.

Como se puede observar para señales en SD se pueden tomar seis medidas por minuto, no así en el caso de la resolución HD. La mayoría del tiempo de ejecución es consumido por los indicadores de *Blur*, *Noise* y *Block loss*, como se puede ver en la tabla 6.7. Estos algoritmos pueden ser mejorados a futuro sobre la base del trabajo realizado por el grupo MOAVI para poder realizar una medida continua para señales HD. Estos indicadores en el sistema SMTVD se pueden tomar cada 20 o 30 segundos, incluso a esta tasa la información recolectada es valiosa y útil para generar nuevos modelos de calidad de video utilizando escenas reales de TVD.

Con el objetivo de comprobar que los indicadores MOAVI puedan ser utilizados para obtener un nuevo modelo NR que mejore el desempeño del modelo paramétrico de estimación de I_c de la ecuación 5.3 se comenzó el análisis al evaluar un subconjunto de los indicadores presentados en la tabla 6.5 sobre el conjunto de videos del *Set I HD* y *Set II SD* sin pérdida de paquetes. Se hizo foco en los siguientes indicadores: *Blockiness*, *Blur*, *Brightness*, *Contrast*, *Exposure*, *Flickering*, *Spatial Activity* y *Temporal Activity*.

Tabla 6.5: Indicadores MOAVI utilizados en el sistema de monitorización

Indicador	Descripción
<i>Blockiness</i>	Este indicador mide uno de los fenómenos más reconocidos en el área de compresión de imágenes, el denominado efecto de bloques. Dicho efecto es reconocido en compresión JPEG por el truncamiento de altas frecuencias en los bloques de la DCT
<i>Blur</i>	Indicador de borrosidad, mide el nivel de borroneo de la imagen. Clásico en compresión JPEG 2000 (Wavelets).
<i>Brightness</i>	Indicador de brillo medio
<i>Contrast</i>	Indicador de contraste
<i>Exposure</i>	Este indicador mide la distorsión por tiempo de exposición de la toma.
<i>Flickering</i>	Indicador de parpadeo debido a efectos de compresión. Se basa en el concepto de que al disminuir el bitrate lo que sucede es que los macrobloques cuadro a cuadro no se actualizan si la suma de diferencias absolutas no supera cierto umbral. Esto provoca que haya cambios bruscos en tomas con zoom o paneo que producen el efecto de “parpadeo” local.
<i>SA: Spatial Activity</i>	Actividad espacial, refiere a la complejidad de cada cuadro, relacionado con la cantidad de bordes.
<i>TA: Temporal Activity</i>	Actividad temporal, refiere al movimiento cuadro a cuadro.
<i>Interlace</i>	Este indicador detecta el grado del efecto de entrelazado en un video.
<i>Noise</i>	Indicador de ruido, se computa calculando la varianza local en las regiones de menor actividad espacial en el cuadro o imagen.
<i>Framing</i>	Este indicador detecta si existe encuadramiento de la imagen.
<i>Blackout</i>	Este indicador detecta si hay cuadros negros.
<i>Pillarboxing</i>	Este indicador detecta si se realiza un recuadro de tipo <i>Pillarbox</i> en una secuencia de cuadros
<i>Letterboxing</i>	Este indicador detecta si se realiza un recuadro de tipo <i>Letterbox</i> en una secuencia de cuadros
<i>Slicing</i>	Este indicador detecta si existen <i>slices</i> o tirillas dañadas en una secuencia de cuadros.
<i>Block loss</i>	Encuentra los bordes verticales y horizontales de una imagen y estima el número de macrobloques que fueron perdidos.

Capítulo 6. Evaluación de calidad de video en un sistema de monitorización de la señal de TV Digital

Tabla 6.6: Tiempo de ejecución de I_c y de indicadores MOAVI

Indicador	HD	SD
<i>Indicadores MOAVI</i>	13.5 seg	2.9 seg

Tabla 6.7: % tiempo

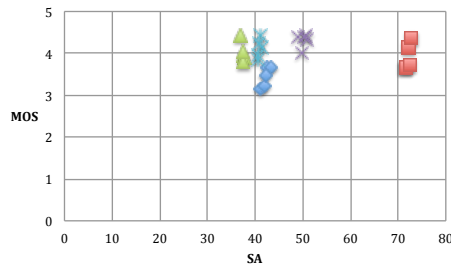
Indicadores	% tiempo de ejecución
Blur	42.57
Noise	15.86
BlockLoss	10.54
Slice	8.30
Flickering	5.18
SpatialAct	4.83
Exposure	3.90
Interlace	3.83
Temporal	1.78
Blockiness	1.59
conBriBlack	1.55
Letterbox	0.0003
Freezing	0.0003
Pillarbox	0.0002

En la figura 6.4 se muestran los resultados obtenidos por los indicadores MOAVI seleccionados sobre los videos en resolución HD del *Set I HD*. Cabe destacar que en [90] estos indicadores fueron evaluados sobre clips con resolución estándar (SD). En estas ilustraciones se puede visualizar la relación de cada uno de estos indicadores con el MOS obtenido para cada clip. En tanto cada punto representa una secuencia de video y las formas de los puntos (cruces, rombos, etc.) corresponden a los diferentes contenidos codificados a su vez a diferentes *bitrates*.

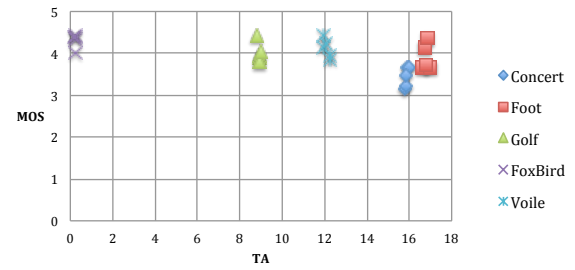
Se puede observar en los resultados obtenidos para los indicadores *Spatial Activity* y *Temporal Activity* que los puntos se ubican en forma “casi vertical”. Esto significa que la calidad subjetiva obtenida en estos clips no presenta clara correlación con estos indicadores sino que estos varían según el contenido del clip. En el caso del indicador de contraste se puede ver que para un mismo contenido el cambio de bitrate no afecta el contraste de la imagen de manera significativa no siendo este fenómeno tan claro en el caso de los indicadores de exposición de la escena y brillo medio. Cabe destacar que en el último caso no se ve una correlación clara con el MOS.

Por otro lado para los indicadores Blockiness y Flicker, la ubicación de los puntos se aleja del patrón de líneas verticales, presentando variaciones apreciables del MOS para cada uno de los contenidos. Esto lleva a pensar que la calidad puede ser evaluada en función de estos últimos indicadores y además en función del contenido del video. El contenido, por su parte, puede ser caracterizado por

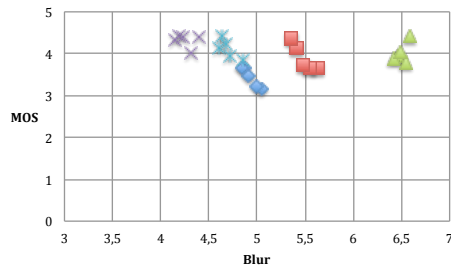
6.7. Indicadores MOAVI



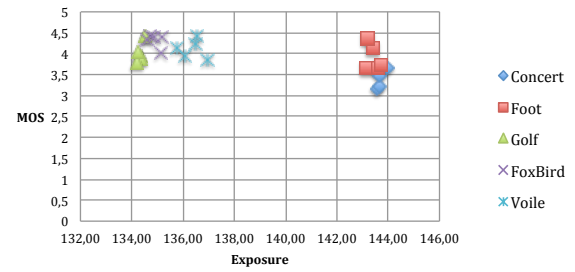
(a) Actividad Espacial



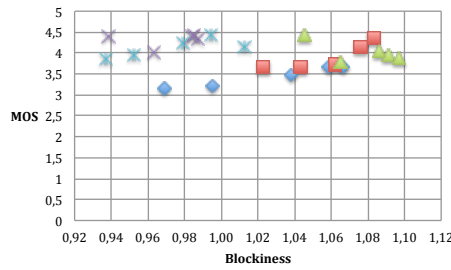
(b) Actividad Temporal



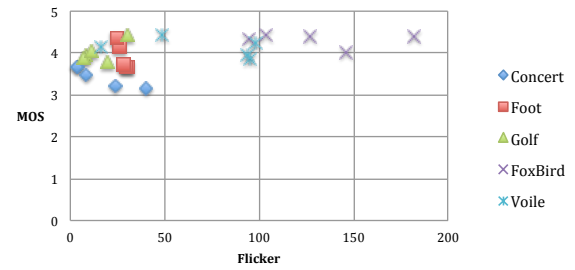
(d) Borroneo



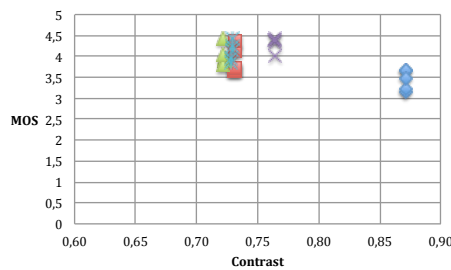
(e) Exposición



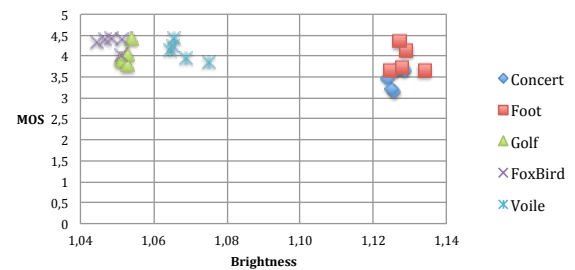
(g) Efecto de Bloques



(h) Parpadeo



(j) Contraste



(k) Brillo

Figura 6.4: Ilustración de resultados de los indicadores MOAVI obtenidos para el conjunto reducido de videos del *Set I HD* con degradaciones debido a artefactos en compresión

Capítulo 6. Evaluación de calidad de video en un sistema de monitorización de la señal de TV Digital

uno o varios de los otros indicadores como el de actividad espacial o temporal.

En una primera aproximación, se utilizó una técnica de aprendizaje supervisado denominada Regresión por Vectores de Soporte (SVR) [91] [92] con el fin de utilizar el resultado de estos indicadores como características y así obtener una estimación del MOS por medio de una regresión. Al utilizar todos los indicadores como entradas para el modelo SVR se pudo comprobar que se obtenía un sobreajuste de los datos en entrenamiento, lo que lleva a tener muy buenos resultados con estos datos, pero malos resultados en el conjunto de validación. El conjunto de datos de validación utilizado se conformo por un subconjunto de videos del *Set II HD* sin pérdida de paquetes. Este resultado fue de esperarse dado que la cantidad de características tomadas para cada video resulta demasiado grande para la cantidad de videos en el conjunto de entrenamiento y validación. Luego de diversos análisis y pruebas reduciendo el número de características y alterando su combinación, se descartó esta aproximación al problema, ya que los resultados obtenidos no fueron suficientemente buenos; en particular, se puede decir que no se logró un nivel mínimo aceptable en términos de correlación con el MOS, especialmente en la etapa de validación. En el conjunto de videos de entrenamiento se consideraron únicamente cinco contenidos diferentes para obtener de forma controlada un conjunto de videos codificados con diferentes bitrates. En todos los casos, el material original era de “excelente calidad”, y las degradaciones de codificación introducidas fueron “menores” en el sentido de que en general el MOS obtenido para estos clips fue mayor a 3 (Aceptable). Esto limita la posibilidad de evaluar exhaustivamente la relación entre varios de los indicadores y la calidad subjetiva e hizo difícil obtener resultados satisfactorios al utilizar métodos de aprendizaje supervisado como SVR al no contar con el grado de variabilidad adecuado para estos indicadores.

Cabe destacar que el bitrate que se utilizó para codificar los videos en la base de datos VQI se encuentra en el rango habitual utilizado en transmisiones en vivo de TV Digital, que es superior al utilizado para *streaming* sobre redes IP como es el caso de Internet donde la reducción de calidad debido al bitrate altera de manera significativa la magnitud de estos artefactos, haciendo que estos sean más perceptibles. Por lo tanto en principio con la diversidad de videos calificados en las bases de datos utilizadas no fue posible obtener por medio de estos indicadores un modelo de I_c que tenga resultados aceptables. Igualmente para la primer versión del sistema SMTVD estos indicadores presentados en la tabla 6.5 se medirán en el nodo de referencia y se almacenarán en el sitio central. Es de esperar que en una siguiente versión del sistema, éstos indicadores puedan utilizarse para obtener mejores estimaciones de I_c al comparar su desempeño sobre una base mas grande de videos adquiridos sobre grabaciones reales de TV Digital y por tanto, comparar su desempeño con el MOS obtenidos en nuevas evaluaciones subjetivas; lo que queda como trabajo a futuro de esta tesis.

Capítulo 7

Conclusiones

En esta tesis se profundiza en el estado del arte de sistemas de evaluación objetiva y subjetiva de calidad de video, en el estándar ISDB-T de Televisión Digital Terrestre Abierta (norma de TVD adoptada por Uruguay), y en particular, en la aplicación de técnicas de medición de calidad de video en TVD.

El desarrollo de la TVD en Uruguay supone una mejora de la calidad de imagen y un uso más eficiente del espectro. Por otro lado favorece al aumento en la diversidad cultural del contenido. La aceptación de esta tecnología juega un papel clave y en gran medida depende de la calidad de experiencia (QoE, *Quality of Experience*) percibida por parte del público. Poder contar con herramientas que permitan obtener esta información brindaría a los operadores, en este caso los radiodifusores, y reguladores la posibilidad de velar por la calidad de experiencia que se ofrece al público. Cabe destacar que medir la QoE no supone una tarea sencilla, es una medida subjetiva e involucra múltiples factores que afectan la percepción global. Siguiendo esta línea se observó 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.

Esta tesis contiene aportes específicos en el área de evaluación objetiva de calidad de video con aplicación en 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. La implementación de dicho sistema se realizó en el marco del proyecto titulado: “Sistema de Monitorización de la señal de TV Digital” (SMTVD) (fondo ANII-FST-1-2013-1-13436).

Con el fin de obtener un estimador global de calidad de video (MOS, *Mean Opinion Score*) para la señal de TVD se parte de un modelo paramétrico NR inicial, propuesto previamente en el proyecto titulado: “Indicadores de Calidad de Video” (VQI, *Video Quality Indicators*) (fondo ANII-FST-1-2012-1-8147). En este trabajo se extiende su funcionamiento tomando en consideración la arquitectura

Capítulo 7. Conclusiones

distribuida del sistema SMTVD. 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 (pérdida de paquetes). 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.

Con el fin de estimar la degradación de calidad percibida debido a pérdida de información en el canal de transmisión se evalúan y analizan distintos modelos objetivos paramétricos NR y modelos FR/RR del estado del arte. Dentro de estos últimos se utilizan los siguientes modelos: VQM *LowBw*, VQM *General*, PSNR, PSNR-HVS, VIFp, MS-SSIM y SSIM. Se muestra que al utilizar modelos FR/RR que comparan ambas señales, de referencia y degradada, adquiridas en el nodo de referencia y remoto, respectivamente, se obtiene una mejora estadística significativa en la estimación de la degradación perceptual de calidad debido a la pérdida de paquetes TS (*Transport Stream Packets*) en comparación con los modelos paramétricos NR inicialmente propuestos. Con el fin de seleccionar un modelo a ser utilizado por el sistema para realizar la comparación entre las señales de referencia y degradada se tuvo en cuenta el desempeño de los modelos evaluados por medio de los resultados obtenidos en evaluaciones subjetivas y su rendimiento en término de tiempo de ejecución. 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 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 del contenido del clip de video, denominado SAD. Cabe destacar que se utiliza un modelo NR dado que en este caso solo se cuenta con la señal recibida (no se tiene acceso a la señal original). Con estas dos medidas el sistema es capaz de realizar una estimación global de la calidad en cada nodo de medición.

Para realizar las comparaciones en desempeño entre los modelos propuestos se utilizó un conjunto de clips de video etiquetados con sus correspondientes calificaciones de calidad (MOS) obtenidas por medio de la realización de evaluaciones subjetivas. En el proceso de implementación de estas pruebas se obtuvo un to-

tal de 8900 calificaciones individuales; como resultado la base de datos generada contiene un total de 400 clips de video etiquetados con su respectivo MOS. Entre estos se pueden encontrar secuencias grabadas de aire, secuencias de laboratorio, con distintos esquemas de codificación de video, tasas de bitrate, resolución (HD o SD) y distintos patrones de pérdida de paquetes TS. La base de datos generada en conjunto con las herramientas desarrolladas para su ejecución se encuentran disponibles para la comunidad científica [78].

Como complemento dentro del sistema de monitorización se computan indicadores desarrollados en el proyecto MOAVI que proveen información valiosa sobre la señal monitorizada y su recolección permite obtener información estadística que puede ser utilizada para mejorar los modelos de estimación de calidad empleados, en particular sobre la señal registrada en el nodo de referencia.

Todos los objetivos específicos planteados en esta tesis, descritos en la sección 1.3, fueron cumplidos. Se realizó un estudio del estado del arte en modelos objetivos para evaluación de calidad de video con referencia completa (FR), de referencia reducida (RR) y sin referencia (NR). Se trabajó con herramientas de extracción de características sobre clips de video, es el caso del popular software *FFmpeg*, *VirtualDub*, entre otros. Se tuvo activa participación en el desarrollo de evaluaciones subjetivas de calidad y en el análisis de sus resultados. Se analizaron las degradaciones de calidad asociadas al proceso de compresión de video (compresión H.264/AVC) y al proceso de transmisión de la señal de TVD. Se evaluaron distintos modelos FR/RR y se seleccionó uno de ellos para el cálculo de las degradaciones asociadas al proceso de transmisión. Por otro lado se analizó y seleccionó un modelo NR para el cálculo de las degradaciones asociadas al proceso de adquisición y codificación. Por último, con el fin de integrar el modelo FR seleccionado para su cómputo en tiempo real al sistema SMTVD se debió tener en cuenta la alineación temporal de las secuencias registradas en los distintos nodos del sistema. Específicamente, para solucionar este problema se generó un algoritmo capaz de alinear dichas secuencias (de referencia y degradada) y se modificó el código de la herramienta VQMT, responsable de correr el modelo SSIM, para adecuarla a los requisitos del sistema SMTVD.

Como líneas de investigación a futuro con el sistema de SMTVD funcionando y recolectando grabaciones en distintas ubicaciones se considera pertinente realizar una nueva tanda de evaluaciones subjetivas de videos grabados de aire bajo el nuevo paradigma de nodo de referencia y nodo remoto. En este caso se recolectaría un abanico más amplio de clips con mayor diversidad de contenido, distintos esquemas de codificación, patrones de pérdida de paquetes, permitiendo un mayor análisis, no solo para ampliar y validar el espectro de aplicación de los modelos de calidad propuestos sino para mejorar el desempeño de la estimación de calidad, particularmente la estimación de la degradación perceptual de la señal en el nodo de referencia. Por ejemplo, por medio de la inclusión de los indicadores MOAVI a los modelos NR de estimación de calidad en este punto.

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

Capítulo 7. Conclusiones

transmitido. Si bien en este trabajo el foco de aplicación de los modelos de calidad ha sido en TVD, 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).

Por último, resulta interesante abordar la diversidad que subyace en el proceso de puntuación subjetiva de calidad, permitiendo un análisis más detallado sobre el grado de fiabilidad de la predicción objetiva en términos de variabilidad, por medio de la generación de indicadores que permitan obtener un mejor entendimiento sobre la aplicación o servicio que se está brindando.

Apéndice A

Publicaciones realizadas en el marco de este trabajo

J.P. Garella, E. Grampín, R. Sotelo, J. Baliosian, J. Joskowicz, G. Guimerans, M. Simon, “Monitoring QoE on digital terrestrial TV: A comprehensive approach,” *In IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Nara, Japon, 2016, pp. 1-6.

J. Joskowicz, R. Sotelo, J. P. Garella, P. Zinemanas, M. Simón, “Combining Full Reference and No Reference Models for Broadcast Digital TV Quality Monitoring in Real Time,” *In IEEE Transactions on Broadcasting* , vol.PP, no.99, pp.1-15

J. Joskowicz, R. Sotelo , D. Durán, M. Juayek, J. P. Garella, “Subjective Video Quality Test: Methodology, Database and Experience,” *In IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Ghent, Belgium, 2015.

D. C. Mocanu, J. Pokhrel, J. P. Garella, J. Seppänen, E. Liotou, and M. Narwaria. “No-reference video quality measurement: added value of machine learning,” *Journal of Electronic Imaging*, 24(6):061208, 2015. Disponible en línea: <http://dx.doi.org/10.1117/1.JEI.24.6.061208>.

P. F. Guridi , G. Guimerans, J. P. Garella, J. Baliosian, E. Grampín, R. Sotelo, M. Simon, “Development of a Digital TV Receivers Test Suite” , *Revista Dyna*, Universidad Nacional de Colombia, Medellín, Colombia, 2015.

Nassi, M., Guimerans, G., Guridi, P. F., Garella, J. P., Baliosian, J., Grampín, E., Simon, M. “Interactivity in ISDB-Tb Digital Television: A yet-missing standard”. *In VI International Conference on Interactive Digital TV IV Iberoamerican Conference on Applications and Usability of Interactive TV* (p. 327). 2015.

D. Durán, J. P. Garella, M. Juayek, J. Joskowicz, R. Sotelo, “Estimación automática de calidad de video en televisión digital abierta ISDB-T basada en indi-

Apéndice A. Publicaciones realizadas en el marco de este trabajo

cadoreos objetivos y subjetivos”. *Memoria Investigaciones en Ingeniería*, Volume 12, pp 7-16, 2014.

J. Joskowicz, R. Sotelo , D. Durán, M. Juayek, J. P. Garella, “Automation of Subjective Video Quality Measurements,” In *Proceedings of the Latin America Networking Conference on LANC 2014*, page 7, Montevideo, Uruguay, Sept. 2014.

P. F. Guridi, J. P. Garella, G. Guimerans, J. Baliosian, E. Grampín, M. Simon, “Testing ISDB-Tb Digital TV Receivers”, In *XXI Congreso Internacional de Ingeniería Electrónica, Eléctrica y Computación*, Arequipa, Perú, Aug. 2014.

Referencias

- [1] Patrick Le Callet, Sebastian Möller, Andrew Perkiš, 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] D. C. Mocanu, J. Pokhrel, J. P. Garella, J. Seppänen, E. Liotou, and M. Narwaria. No-reference video quality measurement: added value of machine learning. *Journal of Electronic Imaging*, 24(6):061208, 2015. Disponible en <http://dx.doi.org/10.1117/1.JEI.24.6.061208>, doi: 10.1117/1.JEI.24.6.061208.
- [4] Video Quality Experts Group (VQEG). Disponible en <http://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx>.
- [5] Recommendation ITU-R BT.500-13. Methodology for the subjective assessment of the quality of television pictures. Jan. 2012.
- [6] Recommendation ITU-R BT.710-4. Subjective assessment methods for image quality in high definition television. Nov. 1998.
- [7] Recommendation ITU-T P.910. Subjective video quality assessment methods. Apr. 2008.
- [8] 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.
- [9] B. P. van den Berge. Influence of level of motion on perceived video quality in mpeg-2 coding. Master's thesis, Faculty of Industrial Engineering and Innovation Sciences, Human Technology Interaction, Eindhoven University of Technology, 2009.
- [10] Q. Huynh-Thu, M. N. Garcia, F. Speranza, P. Corriveau, and A. Raake. Study of rating scales for subjective quality assessment of high-definition video. *IEEE Transactions on Broadcasting*, 57(1):1–14, Mar. 2011.

Referencias

- [11] A. Conrad Bovik K. Seshadrinathan, R. Soundararajan and L. K. Cormack. Study of subjective and objective quality assessment of video. *IEEE Transactions on Image Processing*, 19(6):1427–1441, June 2010.
- [12] O. Nemethova, M. Ries, A. Dantcheva, S. Fikar, and M. Rupp. Test equipment of time-variant subjective perceptual video quality in mobile terminals. In *International Conference on Human Computer Interaction*, pages 14–16, Phoenix, USA, Nov. 2005.
- [13] N. Narvekar T. Liu, G. Cash and J. Bloom. Continuous mobile video subjective quality assessment using gaming steering wheel. In *Sixth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, Scottsdale, Arizona.
- [14] M. Pinson. The consumer digital video library [best of the web]. *IEEE Signal Processing Magazine*, 30(4):172–174, 2013.
- [15] 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. Disponible en http://dbq-wiki.multimediatech.cz/_media/qi0306.pdf.
- [16] 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)*, pages 1162–1167, Dec. 2013.
- [17] 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*, 98(1):2–11, 2015.
- [18] 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*, 57(2):165–182, June 2011.
- [19] Y. Chen, K. Wu, and Q. Zhang. From qos to qoe: A tutorial on video quality assessment. *IEEE Communications Surveys and Tutorials*, 17(2):1126–1165, 2015.
- [20] H. Wu and K. R. Rao. *Digital video image quality and perceptual coding*. CRC press, 2005.
- [21] A. Takahashi, D. Hands, and V. Barriac. Standardization activities in the itu for a qoe assessment of iptv. *IEEE Communications Magazine*, 46(2):78–84, Feb. 2008.
- [22] 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*, volume 1, pages 477–480, 2002.

- [23] K. Yamagishi and T. Hayashi. Qrp08-1: Opinion model for estimating video quality of videophone services. In *IEEE Global Telecommunications Conference (Globecom)*, pages 1–5, Nov 2006.
- [24] Recommendation ITU-T G.1070. Opinion model for video-telephony applications. Apr. 2007.
- [25] A. Raake, M. N. Garcia, S. Moller, J. Berger, F. Kling, P. List, J. Johann, and C. Heidemann. T-v-model: Parameter-based prediction of iptv quality. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1149–1152, Mar. 2008.
- [26] M.N. Garcia, R. Schleicher, and A. Raake. Towards a content-based parametric video quality model for iptv. In *Proceedings of the 3rd International Workshop on Perceptual Quality of Systems (PQS'10)*, pages 20–25, 2010.
- [27] A. Khan, L. Sun, and E. Ifeakor. Content-based video quality prediction for mpeg4 video streaming over wireless networks. *Journal of Multimedia*, 4(4):228–239, 2009.
- [28] A. Khan, L. Sun, and E. Ifeakor. Qoe prediction model and its application in video quality adaptation over umts networks. *IEEE Transactions on Multimedia*, 14(2):431–442, Apr. 2012.
- [29] J. Joskowicz and J.C.L Ardao. Combining the effects of frame rate, bit rate, display size and video content in a parametric video quality model. In *Proceedings of the 6th Latin America Networking Conference*, pages 4–11, Quito, Ecuador, Oct. 2011.
- [30] J. Joskowicz, R. Sotelo, and J. C. L. Ardao. Towards a general parametric model for perceptual video quality estimation. *IEEE Transactions on Broadcasting*, 59(4):569–579, Dec. 2013.
- [31] S. Winkler and P. Mohandas. The evolution of video quality measurement: From psnr to hybrid metrics. *IEEE Transactions on Broadcasting*, 54(3):660–668, Sept. 2008.
- [32] 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*, volume 4, Scottsdale, USA, 2006.
- [33] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and Vladimir Lukin. On between-coefficient contrast masking of dct basis functions. In *Proceedings of the third international workshop on video processing and quality metrics*, volume 4, 2007.
- [34] 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*, 13(4):600–612, Apr. 2004.

Referencias

- [35] Z. Wang and X. Shang. Spatial pooling strategies for perceptual image quality assessment. In *IEEE International Conference on Image Processing*, pages 2945–2948, Oct. 2006.
- [36] 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*, volume 2, pages 1398–1402, Nov. 2003.
- [37] Z. Wang and Q. Li. Information content weighting for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 20(5):1185–1198, May. 2011.
- [38] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, Feb. 2006.
- [39] 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*, 18(1):17–33, 2003.
- [40] 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*, 11(1):89–123, 2001.
- [41] M. H. Pinson and S. Wolf. A new standardized method for objectively measuring video quality. *IEEE Transactions on Broadcasting*, 50(3):312–322, Sept. 2004.
- [42] Recommendation ITU-T J.144. Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference. 2004.
- [43] 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.
- [44] Recommendation ITU-T BT.601-7. Studio encoding parameters of digital television for standard 4:3 and widescreen 16:9 aspect ratios. 2011.
- [45] 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*, pages 23–25, 2005.
- [46] VQEG. Validation of Reduced-Reference and No-Reference Objective Models for Standard Definition Television, Phase I. 2009.
- [47] Recommendation ITU-T J.249. Perceptual video quality measurement techniques for digital cable television in the presence of a reduced reference. 2010.

- [48] 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*, pages 9–80, 2000.
- [49] Recommendation ITU-T J.247. Objective perceptual multimedia video quality measurement in the presence of a full reference. 2008.
- [50] Recommendation ITU-T J.246. Perceptual visual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference. 2008.
- [51] Recommendation ITU-R BT. 1866. Objective perceptual visual quality measurement techniques for broadcasting applications using low definition television in the presence of a full reference signal. 2010.
- [52] Recommendation ITU-R BT.1867. Objective perceptual visual quality measurement techniques for broadcasting applications using low definition television in the presence of a reduced bandwidth reference. 2010.
- [53] Recommendation ITU-T J.340. Reference algorithm for computing peak signal to noise ratio (psnr) of a processed video sequence with constant spatial shifts and a constant delay. 2010.
- [54] Recommendation ITU-T J.342. Objective multimedia video quality measurement of hdtv for digital cable television in the presence of a reduced reference signal. 2011.
- [55] Recommendation ITU-T J.341. Objective perceptual multimedia video quality measurement of hdtv for digital cable television in the presence of a full reference. 2011.
- [56] Recommendation ITU-T J.343. Hybrid perceptual bitstream models for objective video quality measurements. 2014.
- [57] 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)*, pages 2301–2305, 2014.
- [58] 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*, pages 1–23, 2014.
- [59] 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.
- [60] 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.

Referencias

- [61] Recommendation ITU-R BT.1908. Objective video quality measurement techniques for broadcasting applications using hdtv in the presence of a reduced reference signal. 2012.
- [62] VQEG. Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, Phase II (FR-TV2). 2003.
- [63] N. O. Pisciotta, C. G. Liendo, and R. C. Lauro. *Transmisión de televisión digital terrestre en la norma ISDB-Tb*. Cengage Learning Editores SA de CV, 2013.
- [64] Asociación Brasileña de Normas Técnicas. *ABNT NBR 15601: Televisión Digital terrestre - Sistema de Transmisión*. 2007.
- [65] Asociación Brasileña de Normas Técnicas. *ABNT NBR 15604: Televisión Digital terrestre - Receptores*. 2007.
- [66] Documento de Armonización Parte 1: Hardware. *Foro Internacional de ISDB-T*, 2012. Disponible en http://archivo.presidencia.gub.uy/sci/decretos/2013/05/miem_850_anexo1.pdf.
- [67] ITU-T Recommendation H.264. *Advanced video coding for generic audiovisual services*. 2005.
- [68] Standard ISO/IEC 14496-10. *Information technology – Coding of audio-visual objects – Part 10: Advanced Video Coding*. 2005.
- [69] Asociación Brasileña de Normas Técnicas. *ABNT NBR 15602-1: Televisión digital terrestre - Codificación de video, audio y multiplexación - Parte 1: Codificación de video*. 2007.
- [70] Standard ISO/IEC 14496-3. *Information technology – Coding of audio-visual objects – Part 3: Audio*. 2005.
- [71] Asociación Brasileña de Normas Técnicas. *ABNT NBR 15602-1: Televisión digital terrestre - Codificación de video, audio y multiplexación - Parte 2: Codificación de audio*. 2007.
- [72] Recommendation ITU-R BT.601. Parámetros de codificación de televisión digital para estudios con formatos de imagen normal 4:3 y de pantalla ancha 16:9.
- [73] Protocolo de Homologación de Receptores ISDB-Tb, [online], <http://www.dinatel.gub.uy/-/publican-protocolo-de-homologacion-de-receptores-de-tvd>. Accessed: 21 September 2016.
- [74] Asociación Brasileña de Normas Técnicas. *ABNT NBR 15603-1: Televisión digital terrestre - Multiplexación y servicios de información (SI) - Parte 1: Servicios de información del sistema de radiodifusión*. 2007.

- [75] Asociación Brasileña de Normas Técnicas. *ABNT NBR 15602-3: Televisión digital terrestre - Codificación de video, audio y multiplexación - Parte 3: Sistemas de Multiplexación de Señales*. 2007.
- [76] O. Edfors, M. Sandell, J.-J. van de Beek, D. Landström, and F. Sjöberg. An introduction to orthogonal frequency-division multiplexing. *Div. of Signal Processing, Research Report*, 1996.
- [77] Benigno Rodríguez Díaz. *Differential STBC for OFDM Based Wireless Systems*. Cuvillier Verlag, 2007.
- [78] 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*, pages 1–6, June 2015.
- [79] 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*, page 7. ACM, 2014.
- [80] 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, IM-QA2008*, page 6, Kyoto, Japan, September 2008. Disponible en <https://hal.archives-ouvertes.fr/hal-00300182>.
- [81] FFmpeg, [online], <http://www.ffmpeg.org>.
- [82] 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*, 2014:1–11, 2014. Disponible en <http://dx.doi.org/10.1155/2014/242531>.
- [83] E. Karapanos, J.-B. Martens, and M. Hassenzahl. Accounting for diversity in subjective judgments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 639–648. ACM, 2009.
- [84] T. Hobfeld, R. Schatz, and S. Egger. Sos: The mos is not enough! In *Third International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 131–136, Sept. 2011.
- [85] J. Joskowicz. *Desarrollo de un Modelo Paramétrico General de Estimación de la Calidad Percibida de Video*. PhD thesis, Ingeniería Telemática, Universidad de Vigo, Oct. 2012.
- [86] S. Winkler. Analysis of public image and video databases for quality assessment. *IEEE Journal of Selected Topics in Signal Processing*, 6(6):616–625, Oct. 2012.

Referencias

- [87] 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*, pages 407–418, 2006.
- [88] 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*, pages 1–4, June 2014.
- [89] VQEG Final Draft Report, Validation of objective models of video quality assessment, Phase II, Mar. 2003 .
- [90] P. Romaniak. *Assessment of Perceptual Video Quality Affected by Acquisition and Bit-rate Reduction Artifacts*. PhD thesis, AGH University of Science and Technology Faculty of Electrical Engineering, Automatics, Computer Science and Electronics Department of Telecommunications, Al. Mickiewicza 30, 30-059 Kraków, Poland, 2011.
- [91] A. J. Smola and V. Vapnik. Support vector regression machines. *Advances in neural information processing systems*, 9:155–161, 1997.
- [92] A. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.

Índice de tablas

2.1.	Escala discreta de cinco niveles, (DSIS)	9
2.2.	Escala discreta de cinco niveles, (ACR)	10
3.1.	Asignación de valores PID	39
4.1.	Secuencias originales seleccionadas para el <i>Set I HD</i> & <i>Set SD</i> . .	49
4.2.	Parámetros de codificación H.264	50
4.3.	Patrones de pérdida de paquetes TS probados	52
4.4.	Condiciones generales de observación para efectuar evaluaciones subjetivas en el entorno de laboratorio	55
4.5.	Sesiones de pruebas subjetivas	56
5.1.	Valores de parámetro: a	68
5.2.	Parámetros del modelo paramétrico: $c1 - c6$	69
5.3.	Valores para los coeficientes $k, x1, x2$	70
5.4.	Valores para los coeficientes α, β	72
5.5.	Comparación de los resultados obtenidos para los modelos de esti- mación de I_p propuestos	73
5.6.	Comparación de los resultados obtenidos en entrenamiento y vali- dación de los modelos seleccionados	74
6.1.	Tiempo de ejecución del modelo asociado a I_c	83
6.2.	Desempeño de los modelos FR/RR propuestos sobre los videos con pérdida de paquetes ($I_p < 1$) de los conjuntos <i>Set I HD</i> y <i>Set I SD</i> , 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)	88
6.3.	Desempeño de los modelos FR/RR/NR propuestos sobre los clips de video con pérdida de paquetes TS ($I_p < 1$) de los conjuntos <i>Set I HD</i> y <i>Set I SD</i> , se utiliza la PLCC (Correlación de Pearson) y RMSE (Raíz del Error Cuadrático Medio).	92
6.4.	Proceso temporal de alineación de cuadros	96
6.5.	Indicadores MOAVI utilizados en el sistema de monitorización . . .	99
6.6.	Tiempo de ejecución de I_c y de indicadores MOAVI	100
6.7.	% tiempo	100

Índice de figuras

2.1. Ejemplo de escalas verticales impresas para evaluación subjetiva de la calidad mediante el método DSCQS.	9
2.2. Visibilidad de distintos tipos de distorsión, ambas imágenes tienen el mismo PSNR aunque su calidad percibida sea distinta. Imagen extraída de [31].	14
2.3. Imágenes degradadas con igual MSE. Se aplica la métrica MSSIM obteniéndose una mejor correlación con la calidad percibida.	17
2.4. Imagen Original, Imagen degradada, Absolute Error Map, SSIM Index Map. Imagen extraída de [35].	17
2.5. La información mutua entre la señal C y E cuantifica la información que el cerebro idealmente puede 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.	19
2.6. Propiedades del modelo VIF: Imagen de referencia (a), con mejora de contraste (b), con compresión JPEG (c) y con borrosidad por medio de un filtro pasa bajos (d).	20
3.1. Ilustración de la utilización de los distintos sistemas de Televisión Digital por país. Fuente: <i>Wikipedia</i>	29
3.2. Diagrama de bloques de un codificador de video	32
3.3. Diagrama de bloques de un decodificador de video	33
3.4. Patrones de muestreo YUV444, YUV422, YUV420 (progresivos)	35
3.5. Cuadro YUV420p. Ilustración basada en una imagen extraída de Wikipedia.	35
3.6. Diagrama en bloques de un <i>encoder</i> genérico	38
3.7. Estructura del encabezado de un paquetes TS	40
3.8. Captura de pantalla de Encoder utilizado para la generación de un flujo MPEG-2 TS ISDB-T. Equipo: Z3 Technology, LLC MVE-02.	42
3.9. Diagrama en bloques del sistema de transmisión	43
3.10. Tasa de datos de un único segmento. Tabla basada en la norma ABNT NBR 15601 <i>Televisión Digital Terrestre</i> - Sistema de Transmisión	44

Índice de figuras

3.11. Tasa total de datos para 13 segmentos. Tabla basada en la norma ABNT NBR 15601 <i>Televisión Digital Terrestre</i> - Sistema de Transmisión.	45
3.12. Configuración de parámetros de transmisión para una señal ISDB-T en un equipo profesional modulador de TVD. Equipo: EITV Playout Pro.	46
4.1. Vista previa de cuadros afectados por simulación de errores en transmisión en cada uno de los contenidos seleccionados	51
4.2. Vista previa de cuadros sin pérdida de paquetes TS de grabaciones tomadas por aire	53
4.3. Vista previa de cuadros afectados por errores en transmisión en algunas secuencias grabadas de aire	53
4.4. Ambiente controlado de pruebas subjetivas	54
4.5. MOS promedio en función del tipo de Televisor (CRT, LCD o LED) que utilizan los sujetos de prueba	57
4.6. MOS promedio en función de la cantidad de horas que los usuarios miran televisión	58
4.7. Ilustración de resultados de calidad obtenidos para el Set I HD & SD	59
4.8. Desviación Estándar en función del MOS para los conjuntos Set I HD, Set I HD, Set II HD y Set II SD	60
4.9. PDU vs MOS para los conjuntos Set I HD, Set I HD, Set II HD y Set II SD	62
4.10. PDU vs MOS, se toma U con los siguientes valores 2,3,4 y 5 y se aplica para los conjuntos Set I HD, Set I HD, Set II HD y Set II SD	63
5.1. Ilustración del método de calculo de la característica s (SAD) . . .	67
5.2. Estimación de I_c sobre la base de datos de entrenamiento	69
5.3. Ilustración de los resultados en la estimación de I_p al utilizar las ecuaciones 5.10 y 5.11 sobre los videos del conjunto de entrenamiento <i>Set I</i> con I_p menor a la unidad, en (a) se puede observar I_p en función de P_w y en (b) la dispersión entre I_p y su predicción $I_{p_{pred}}$	71
5.4. Ilustración de los resultados en la estimación de I_p , en (a) se presenta el ajuste exponencial, en (b) se tiene la dispersión de I_p en función de su predicción $I_{p_{Exp}}$, en (c) se presenta el ajuste polinómico, en (d) se tiene la dispersión de I_p en función de su predicción $I_{p_{Pol}}$	72
5.5. Resultados obtenidos por el modelo $MOS_{p_{Exp}}$ en los conjuntos (a) <i>Set I</i> y (b) <i>Set II</i> , respectivamente.	74
5.6. Resultados obtenidos por el modelo $MOS_{p_{Pol}}$ en los conjuntos (a) <i>Set I</i> y (b) <i>Set II</i> , respectivamente.	74
6.1. Etapas principales en la cadena de TVD donde es importante realizar un control de calidad	79
6.2. Puntos donde realizar una evaluación objetiva de la calidad de la programación	79

6.3.	Ilustración a alto nivel de la arquitectura del sistema de monitorización en tiempo real de la señal de TVD	81
6.4.	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 <i>Set I HD</i> con (a) distribución uniforme de pérdida de paquetes (b) dos ráfagas (<i>bursts</i>) 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 <i>slice</i> de tipo B	86
6.2.	Desempeño de los modelos PSNR, PSNR-HVS, SSIM, MS-SSIM, VIFp, VQM “General” y VQM “LowBw”	91
6.3.	Dispersión <i>MOS vs MOS_{pSSIM}</i> sobre el conjunto <i>Set I HD</i> y <i>Set I SD</i>	92
6.4.	Ilustración de resultados de los indicadores MOAVI obtenidos para el conjunto reducido de videos del <i>Set I HD</i> con degradaciones debido a artefactos en compresión	101

Compilado el sábado 1 abril, 2017.
<http://iie.fing.edu.uy>