

COMPARACIÓN DE CALIDAD DE VIDEO ENTRE CODIFICADORES HARDWARE DE LOS ESTÁNDARES H.264/AVC Y AVS1

Ing. Humberto Camacho Domínguez¹, Ing. José Ángel Espinosa Roque²

¹CUJAE, Instituto Superior Politécnico José Antonio Echeverría, Calle 114, # 11901 e/ Ciclovía y Rotonda,
Marianao, La Habana, Cuba,

²LACETEL, Instituto de Investigación y Desarrollo de Telecomunicaciones, Ave. Independencia, N° 34515, Km 14
½, Reparto Iro de Mayo, Municipio Boyeros, La Habana, Cuba.

¹e-mail: sylar9107@gmail.com

²e-mail: jose@lacetel.cu

RESUMEN

Actualmente en Cuba está aprobado el empleo de tres estándares de codificación de video: MPEG 2, H.264/AVC y AVS1 para el despliegue de la Televisión Digital. El siguiente trabajo forma parte de las investigaciones que se están realizando en **LACETEL**, Instituto de Investigación y Desarrollo en Telecomunicaciones, las cuales están relacionadas al estudio y análisis de sistemas alternativos de codificación de segunda generación utilizados en la Televisión Digital (ej. AVS1). Se presentan los detalles metodológicos, resultados y experiencias de una evaluación de calidad subjetiva entre la eficiencia de codificación de implementaciones de los estándares de compresión de video AVS1 y H.264/AVC, para videos de definición estándar en perfil Jizhun y Principal respectivamente, acordes a la recomendación de la Unión Internacional de Telecomunicaciones apropiada, empleando para ello 5 contenidos de video y 3 velocidades de bits, con el fin de validar similitudes entre estos dos estándares.

PALABRAS CLAVE: AVS1, H.264/AVC, codificación, estándar, subjetiva.

ABSTRACT

Nowadays in Cuba there are three codification standards approved for Digital Television deployment: MPEG-2, H.264/AVC and AVS1. This paper is part of investigations that are taking place at **LACETEL**, Research and Development Telecommunications Institute, related with the study and analysis of second generation alternative codification systems used in Digital Television (e.g. AVS1). Methodological details, results and experiences of a subjective video quality assessment of coding efficiency between encoders implementations of AVS1 and H.264 / AVC video compression standards are presented. The configuration profiles employed are Jizhun profile and Main profile respectively for standard definition video, according to the proper International Telecommunications Union recommendation, employing five test videos and 3 bit rates, in order to validate similarities between these two hardware solutions.

KEYWORDS: AVS1, H.264/AVC, codification, standard, subjective.

INTRODUCCIÓN

Durante los últimos 20 años, la comunicación analógica alrededor del mundo se ha visto desplazada por la comunicación digital. El vídeo digital al igual que los contenidos multimedia distribuidos a través de redes de telecomunicaciones se ha convertido en un tema de suma importancia a lo largo de las últimas décadas, así como objeto de mucha transformación con grandes avances debido a su utilización en una amplia gama de aplicaciones, entre ellas, la televisión digital. Con el aumento, de interés comercial, en las comunicaciones, la necesidad de la existencia de estándares internacionales de compresión de imagen y video tuvo lugar, ya que posibilita la interoperación de los productos de diferente manufactura como codificadores, decodificadores y medios de almacenamiento.

El estándar de compresión de video H.264/AVC (Advanced Video Coding), es actualmente el más globalizado y con mayor soporte industrial, por lo que representa uno de los códec más avanzados siendo ampliamente utilizado en una gran variedad de aplicaciones incluyendo la difusión de la señal de TV en HD sobre satélite, cable y sistemas de transmisión terrena, en aplicaciones de seguridad, videos para redes móviles e Internet, así como para otros tipos de aplicaciones como las de video en demanda y sistemas de tele-presencia.[1] Marcó el inicio de una nueva generación dentro de los sistemas de compresión que se venían utilizando en el mundo y con él se abrió paso a una nueva etapa en el desarrollo de la Televisión Digital.[2]

Por otra parte, AVS1 (Audio Video Coding Standard) es el último estándar de codificación de vídeo digital de segunda generación desarrollado por AVS working group of China, producto de la necesidad de reducir los royalties pagados por el pueblo chino para el uso de otras normas internacionales de codificación de vídeo tales como MPEG-2, MPEG-4 y MPEG-4 parte 10 (H.264), así como para promover la electrónica china permitiendo que sus productos lleguen a un mercado de masas, [3]. Es un estándar de alta eficiencia que emplea las últimas herramientas de compresión de video que abordan principalmente la de definición estándar (SD) y de alta definición (HD), destinadas a alcanzar una eficiencia de codificación similar a H.264/AVC, pero con una menor complejidad computacional. AVS1 fue desarrollado con el fin de estar enfocado a las aplicaciones de radiodifusión y de medios de almacenamiento tales como la televisión digital, disco de vídeo digital (DVD y disco de alta definición) y a las aplicaciones multimedia de banda ancha como la videoconferencia, el vídeo bajo demanda, IPTV, etc. [4]

En un sistema de procesamiento y transmisión de video, la señal se encuentra en constante exposición a degradaciones, distorsiones (durante el proceso de adquisición, compresión, procesamiento, transmisión y reproducción), pérdidas de cuadros y otros fenómenos indeseables que afectan la calidad de la imagen percibida por los usuarios. Por esta razón, las técnicas de evaluación de calidad de video desempeñan un papel de suma importancia para los proveedores de servicios y generadores de contenido, ya que les permite un monitoreo y una evaluación de la imagen recibida en el extremo de dicho sistema.

La evaluación de calidad de video puede ser dividida en dos categorías: evaluación de calidad objetiva y subjetiva. A simple vista puede deducirse que los métodos objetivos no son más que métricas o

algoritmos matemáticos encargados de analizar las diferencias físicas entre la imagen original y la modificada, interpretan los resultados en el dominio espacial o frecuencial, con una tendencia marcada de dar resultados lo más cercanamente posible a evaluaciones humanas. Por otra parte, como su propio nombre sugiere, los métodos subjetivos son realizados a través de los humanos, es decir, están basados en la percepción visual humana [5], para determinar la calidad de funcionamiento de los sistemas probados a través de mediciones que anticipan de manera más directa las reacciones de quienes podrían ver dichos sistemas, representando así los métodos más confiables de medición de calidad, al ser resultado de la opinión directa de los observadores. [6]

Para la evaluación subjetiva de calidad de video, la medición puede ser efectuada mediante los métodos definidos en las Recomendaciones UIT-R BT-500-13 y UIT-T P.913, para imágenes en televisión y aplicaciones multimedia respectivamente. La Recomendación UIT-R BT-500-13 detalla los métodos de escala de degradación con doble estímulo (DSIS, del inglés, Double Stimulus Impairment Scale), escala de calidad continua de doble estímulo (DSCQS, del inglés, Double Stimulus Continuous Quality Scale), los métodos de estímulo único, de comparación de estímulos, y así como los de evaluación de calidad continua de estímulo único (SSCQE, del inglés, Single Stimulus Continuous Quality Evaluation) y el de doble estímulo simultáneo para evaluación continua (SDSCE, del inglés, Simultaneous Double Stimulus for Continuous Evaluation). Mientras, la Recomendación UIT-T P.913 detalla el de índice por categorías absolutas (ACR, del inglés, Absolute Category Rating), por categoría de degradación (DCR, del inglés, Degradation Category Rating), equivalente al DSIS y el método de comparación de pares (PC, del inglés, Pair Comparison), también conocido como índice por categoría de comparación (CCR, del inglés, Comparison Category Rating).

METODOLOGÍA DE PRUEBA

Al efectuar evaluaciones subjetivas adecuadas, en primer lugar, es necesario seleccionar entre las distintas opciones disponibles, aquella que se adapte mejor a los objetivos y circunstancias del problema de evaluación inmediato. Teniendo en cuenta los tiempos de codificación de la señal en cada codificador, debido a que el test se realiza en tiempo real, se determinó para este trabajo el método de escala de calidad continua de doble estímulo (DSCQS, por sus siglas en inglés), en su segunda variante de dos repeticiones, descrita en la Recomendación UIT-R BT-500-13, ya que se cuenta con videos de 10 segundos. En el mismo, se observa un par de imágenes dos veces, de forma secuencial y aleatoria, es decir, estímulos A y B, ambas de la misma fuente, pero habiéndose transmitido una por el sistema que se evalúa, y la otra directamente desde la fuente, para luego evaluar la calidad de ambos estímulos en el segundo par, según la Figura 1, siendo la imagen de referencia de total desconocimiento por parte de los evaluadores.

**COMPARACIÓN DE CALIDAD DE VIDEO ENTRE CODIFICADORES HARDWARE DE LOS ESTÁNDARES
H.264/AVC Y AVS1**

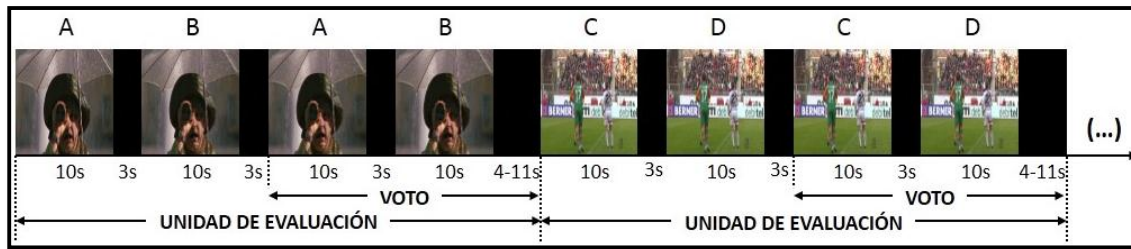


Figura 1: Presentación de los estímulos en el método DSCQS.

La señal o video fuente proporciona directamente la imagen de referencia, y la entrada para el sistema sometido a prueba por lo que deberá ser de calidad óptima para la norma de televisión utilizada. La ausencia de defectos en la parte de referencia del par presentado es esencial para obtener resultados estables [6]. Con el objetivo de abarcar un amplio margen de contenido, se hace necesario la selección de videos de referencia con los cuales se obtenga variada información espacial y temporal. Un total de siete videos sin compresión en formato YUV a 1080p, dos clips para la estabilización de los sujetos de prueba y 5 para la realización del experimento, fueron utilizados como referencia para la conformación de las secuencias, tomados de la Base de Datos de Calidad de Video IRCCyN IVC1080i [7]. Puesto que las pruebas subjetivas eran concebidas para definición estándar, dichos clips fueron previamente convertidos de resolución 1920x1080 píxeles a 720x405 píxeles, respetando la misma relación de aspecto para no sacrificar contenido ni expansión de la imagen, empleando el software libre ffmpeg[8], obteniéndose videos de definición estándar con pocas pérdidas, almacenados como archivos en formato .avi, con muestreo de color YUV 4:2:2, 10 segundos de duración a 25 cuadros por segundo con una relación de aspecto de 16/9, ya que la reproducción de secuencias de video en formato YUV sin compresión, y que a su vez fuese reproducible a pantalla completa, requiere de un software específico exigente de pago (Figura 2).



Figura 2: Captura de clips utilizados.

Para lograr resultados precisos y confiables durante una evaluación subjetiva de calidad de video, se hace necesario un ambiente controlado y un sistema concreto, lo cual resulta de vital importancia para evitar influencias en los resultados de factores externos.

El esquema utilizado para la conformación de la prueba subjetiva es el mostrado en la Figura 3. Los sistemas sometidos a esta evaluación son: codificador H.264/AVC MPEG-4 NOVUS y codificador AVS KuVision, así como el codificador MPEG-2 modelo E202 COMPUNICATE para la señal de referencia.

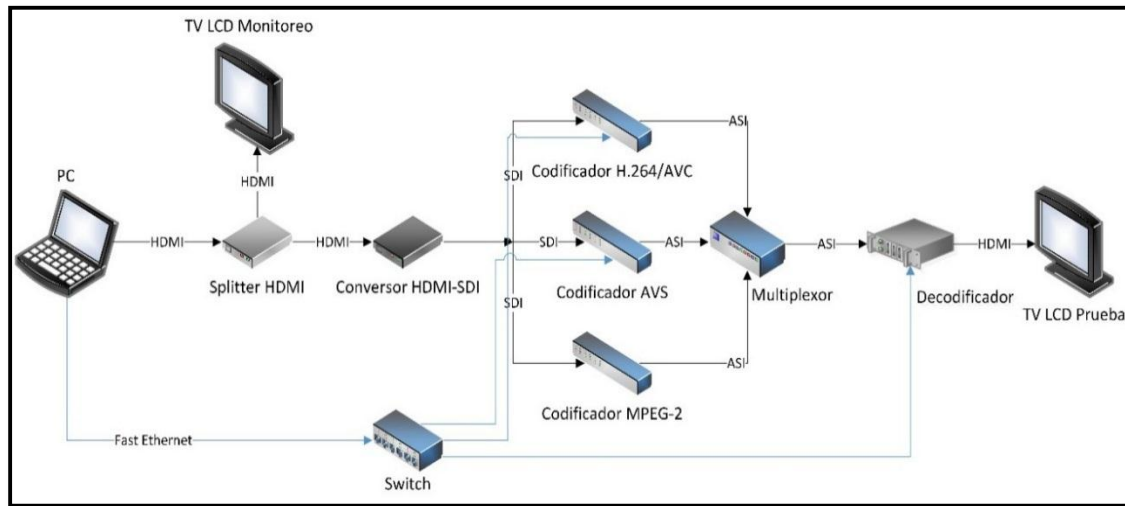


Figura 3: Esquema del equipamiento técnico.

La computadora personal es conectada al sistema a través de la interfaz HDMI (Interfaz Multimedia de Alta Definición, del inglés High-Definition Multimedia Interface) para la reproducción de los distintos clips de video, ya que constituye una norma de audio y vídeo digital cifrado sin compresión. De esta manera se configura dicha interfaz en pantalla extendida a una resolución de 480i, es decir, una resolución de 720 x 480 pixeles, color verdadero (32 bits), a 29Hz., por lo que esta será la resolución de entrada para cada uno de los codificadores. Producto de la necesidad de que el operador visualice el orden y los tiempos de cada secuencia, se hace necesario el uso de un televisor de monitoreo siendo esto posible gracias a un dispositivo denominado Splitter-HDMI para replicar dicha señal. El monitor de referencia utilizado para este fin fue Sony KDL-32BX325 HD de 32 pulgadas. La interfaz de entrada correspondiente a cada uno de los codificadores es SDI (Interfaz Serie Digital, del inglés, Serial Digital Interface), por lo que la señal se convierte en un dispositivo conversor HDMI-SDI, quien entrega a su salida dos flujos SDI. Como la metodología de prueba empleada, en su Recomendación BT.500-13 estipula la evaluación de un sistema de compresión por vez, no existe problema al utilizar una de las salidas del conversor HDMI-SDI para el codificador sometido a prueba y la otra para la señal de referencia. Si la situación amerita el uso de los 3 codificadores al mismo tiempo, el codificador KuVision presenta lazos de flujos SDI en su parte posterior para reutilizar la señal. Una vez codificada la señal de

COMPARACIÓN DE CALIDAD DE VIDEO ENTRE CODIFICADORES HARDWARE DE LOS ESTÁNDARES H.264/AVC Y AVS1

video, cada uno de los flujos ASI convergen en un multiplexor modelo M108C también de marca COMPUNICATE, con el objetivo de conformar un único flujo ASI de información (TS, del inglés, Transport Stream), dirigido a un decodificador profesional KuVision modelo D2000, en donde se permuta o demultiplexa entre una señal codificada u otra para su posterior visualización en un monitor de referencia LG FHD de 43 pulgadas a través de otra interfaz HDMI. Como se ha mencionado anteriormente, dicha prueba se realiza en tiempo real, por lo que los detalles de la configuración y control de los codificadores sometidos a prueba y el decodificador profesional se ejecutan por medio de una red Ethernet de alta velocidad.

Dado a las dimensiones del monitor de prueba, se optó como 4 el número máximo de sillas colocadas para la visualización, puesto que se comprobó que, al adicionar una quinta silla, los sujetos de los extremos sufrían de una percepción menos contrastada que aquellos sujetos dispuestos en el centro de la pantalla. Existen diversas teorías acertadas de distintos organismos acerca de la distancia de observación respecto a las características del monitor, ya sea la altura o la diagonal. Teniendo en cuenta las características del TV empleado y respetando un ángulo horizontal de visión de 30º, la Sociedad de Ingenieros en Televisión e Imágenes en Movimiento (SMPTE, del inglés, Society of Motion Picture and Television Engineer), estipula una distancia (d) de 1.6 veces la diagonal de la pantalla (D), como se muestra en la ecuación (1), para una distancia óptima de visión de 1.75 metros aproximadamente al emplear un TV de 43 pulgadas.

$$d(\text{pulgadas}) = 1.6 \times D(\text{pulgadas}) \quad (1)$$

Los observadores pueden ser expertos o no expertos dependiendo de los objetivos de la evaluación. Un observador experto cuenta con experiencia en las perturbaciones de la imagen que puede introducir el sistema sometido a prueba. Un observador no experto no tiene esta experiencia. En todo caso, los observadores no deben estar directamente familiarizados con el sistema sometido a evaluación, es decir, no deben tener conocimientos específicos y detallados sobre el mismo.

Un total de 20 sujetos no expertos tomaron parte en la campaña de prueba, resultando seleccionados solo 17 debido a que dos de ellos no arrojaron resultados satisfactorios en el test de Ishihara y solo una persona presentaba problemas de agudeza visual para la distancia de visión estimada. Además, luego del análisis de los resultados, ningún sujeto por valores atípicos, fue descartado. Entre los evaluadores figuraba una edad entre 22 y 53 años para un promedio de 29, el 47% de los mismos eran mujeres con un 53% de hombres. Para ver televisión diariamente un 70.59% de los mismos presenta TV de tecnología LCD, mientras que un 47.06% de tecnología CRT ya que existen sujetos que presentan los dos tipos de TV. Pruebas subjetivas de otros autores han arrojado resultados, demostrando la no existencia de diferencias significativas entre las valoraciones de los hombres y las mujeres y tampoco entre las calificaciones por personas de diferentes edades. Sin embargo, puede haber una tendencia de cómo las personas que regularmente ven televisión en monitores de tipo CRT tienden a evaluar mejor la calidad de la imagen en pruebas subjetivas donde se utilizan monitores LCD de elevadas dimensiones, así como una cierta tendencia de las personas que pasan más horas viendo la televisión a dar peores calificaciones de calidad.

Para la selección de las distintas velocidades de bits, se realizó un estudio teórico basado en evaluaciones de calidad de otros autores para videos de definición estándar, el cual oscila entre 700Kb/s y 4Mb/s. Además, se tuvo en consideración las distintas velocidades que se utilizan en el país para la difusión de la señal digital televisiva. De esta forma quedó constituido el uso de 4, 1.5 y 2Mb/s en ese orden para la presentación de las secuencias.

Dos presentaciones consecutivas de dos estímulos más el tiempo de evaluación duran teóricamente 57 segundos. Al emplear 5 contenidos de video útiles más 2 de estabilización, en una sesión se evalúan un total de 17 estímulos para un tiempo aproximado de 16 minutos, es decir, 2 estímulos más 5 contenidos por 3 velocidades de bits, para una evaluación total por sujeto de 34 estímulos, al emplear 2 codificadores. Con el fin de no saturar la prueba se decidió dividir la misma en varios encuentros planificados para 4 días, es decir, 2 días por codificador. El orden de visualización de cada contenido se condicionó de tal forma que existiera variación en cuanto a información espacial y temporal, así como el orden de cada codificación fue aleatorizado.

PROCEDIMIENTO DE PRUEBA.

Una vez seleccionados los sujetos de prueba para una sesión de evaluación de calidad, se les entregó la escala de apreciación la cual contenía un breve formulario donde reflejar su nombre, edad, sexo, el tipo y dimensión del TV que emplea en su casa, además del tiempo que emplea frente a ellos. Seguidamente, previamente ubicados en los distintos puestos, se comienza con la fase de entrenamiento, evitando cualquier tipo de duda respecto a la prueba. Una vez seguros de que cada evaluador se encuentra apto, es decir, se ha familiarizado con el método, comprende el sistema de votación correctamente y presta total concentración y atención a la prueba, se procede a iniciar la sesión de evaluación subjetiva de calidad de video. Un operador situado en la respectiva computadora personal inicializa la reproducción de la secuencia de video mientras que se apoya en el TV de monitoreo para seleccionar entre los distintos estímulos de referencia o codificado, a través de la interfaz del decodificador profesional. Al mismo tiempo, un segundo operador guía a los evaluadores de forma oral por la secuencia que se va transitando, además de servir de referencia al primer operador. Se optó por este mecanismo debido a la tendencia de las personas a desorientarse ya que cada contenido se muestra 4 veces (2 pares). Luego de finalizada la sesión los sujetos entregaron sus escalas de apreciación y dan sus criterios y observaciones respecto a la prueba subjetiva. Desde el momento de entrega hasta la recogida de cada escala, se tomó aproximadamente 20 minutos.

MEDICIONES ESTADÍSTICAS.

Las mediciones estadísticas fueron calculadas para describir, interpretar y analizar la distribución de las anotaciones para cada sujeto en cada una de las condiciones de prueba (combinación de contenido, codificador y velocidad de bits). Para ello se emplea la puntuación de opinión media (MOS, del inglés, Mean Opinion Score) dado por la ecuación (2).

COMPARACIÓN DE CALIDAD DE VIDEO ENTRE CODIFICADORES HARDWARE DE LOS ESTÁNDARES H.264/AVC Y AVS1

$$MOS = \frac{1}{N} \sum_{i=1}^N u_{ijk_r} \quad (2)$$

Donde, u_{ijk_r} representa la nota del observador i para la condición de prueba j , secuencia/imagen k , repetición r , y N , el número de observadores.

RESULTADOS

Nótese que estos valores promedios son calculados sobre la base de la diferencia que cada sujeto de prueba percibió en ambos codificadores respecto a una referencia de video sin codificar, por lo que a menor puntuación de opinión media mejor es la evaluación (Figuras 4,5 y 6).

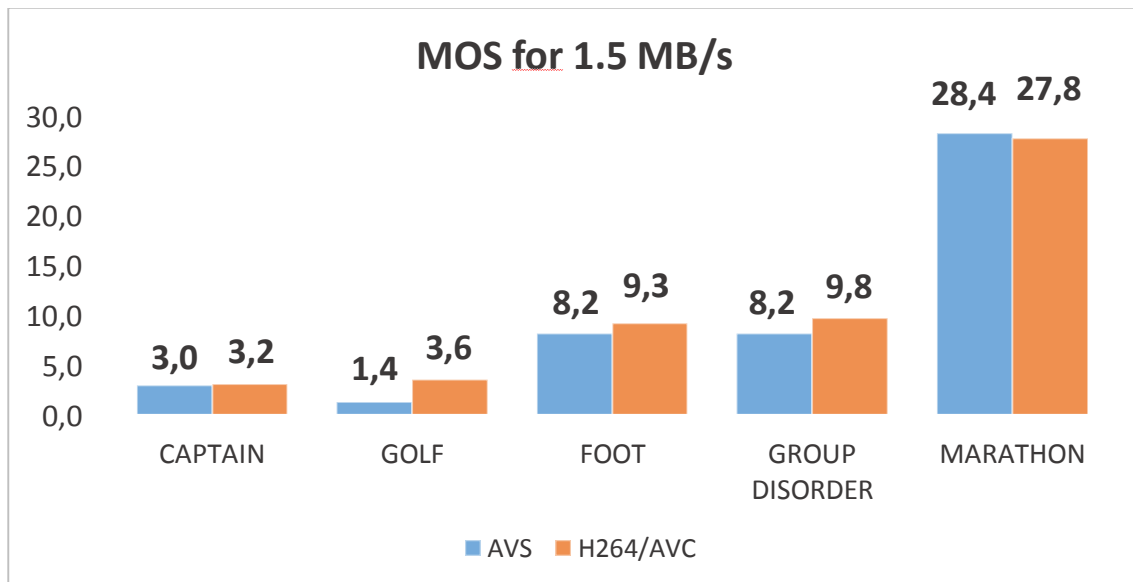


Figura 4. Resultados del MOS para 1.5 MB/s.

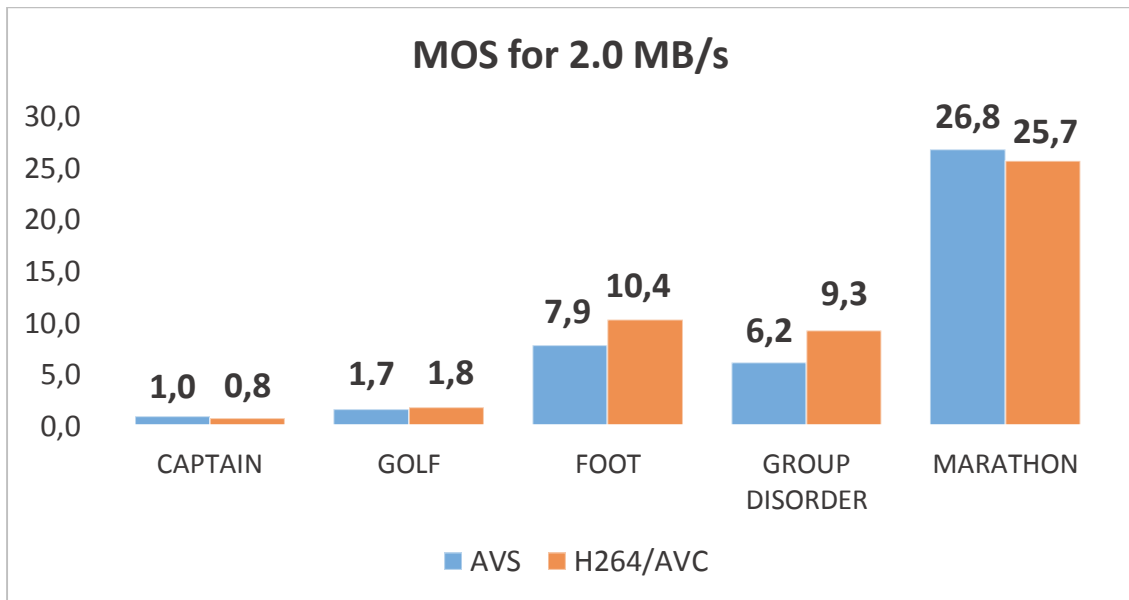


Figura 5. Resultados del MOS para 2.0 MB/s.

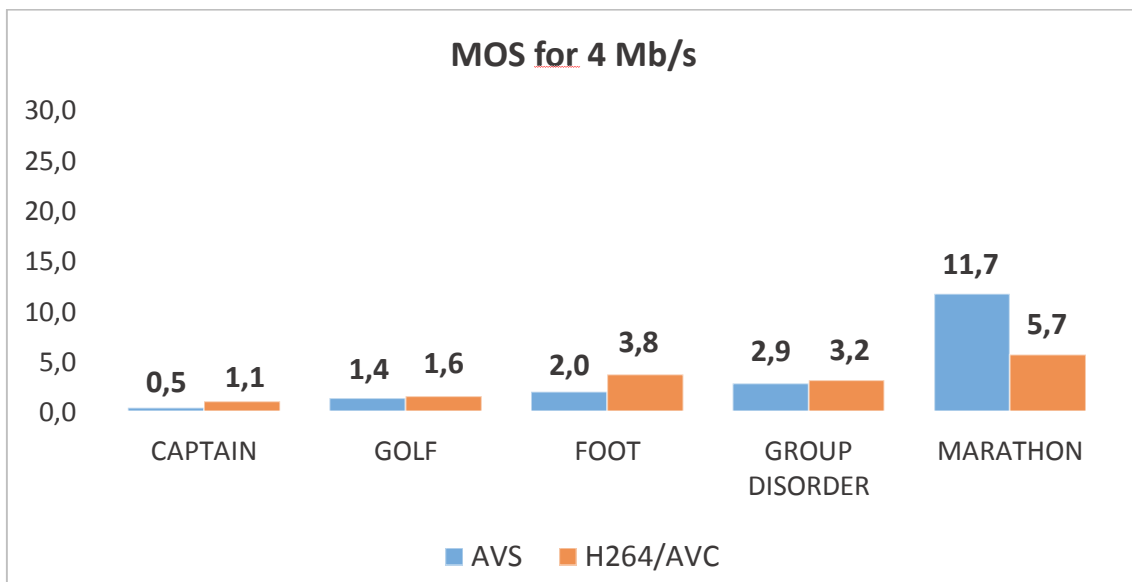


Figura 6. Resultados del MOS para 4.0 MB/s.

CONCLUSIONES.

A lo largo de este proyecto investigativo se ha realizado un análisis y un estudio de cómo llevar a cabo una comparación subjetiva de calidad de video entre codificadores de los estándares de compresión H.264/AVC y AVS1 bajo la metodología expuesta en la Recomendación UIT-R BT-500-13. En cada una de las velocidades de bits empleadas (condiciones de prueba), se pudo observar la similitud existente respecto a eficiencia de codificación de estos dos estándares de compresión de video digital. En videos de poco movimiento (Golf y Captain), gráficamente se pudo observar un comportamiento lineal en respuesta a la variación de las condiciones de prueba, incrementándose ligeramente al reducir a 1.5Mb/s. Ya para videos con un poco más de variación entre cuadros (Group Disorder y Foot), este comportamiento fue variable respecto a la disminución de la velocidad de bit. Para clips , que además de contener considerable información temporal en más de un 50% de su contenido, presentan una elevada información espacial (Above Marathon), se observó un comportamiento de mayor incremento, muy similar, pero logrando el codificador H.264/AVC un mejor procesamiento de imagen que AVS1, traduciéndose en una mejor calidad de video codificado.

REFERENCIAS BIBLIOGRÁFICAS

1. DÍAZ, A.E. AND H.H. LARZÁBAL, Evolución De Los Sistemas De Compresión De Video Más Utilizados En Radiodifusión. 2014, Instituto Superior Politécnico José Antonio Echeverría. Pág. 14-80.
2. RICHARDSON, I.E., The H. 264 advanced video compression standard. 2011: John Wiley & Sons. Pág. 81
3. Audio and Video Coding Standard Workgroup of China. Available from: <http://www.avs.org.cn>.
4. RAO, K., D.N. KIM, AND J.J. HWANG, Video coding standards. The Netherlands: Springer. 2014. Pág. 51-97
5. LIU, T.-J., W. LIN, AND C.-C.J. KUO. Recent developments and future trends in visual quality assessment. In Proceedings of Asia-Pacific Signal and Information Processing Association Annual Submit and Conference. 2011.
6. Assembly, I.R., Methodology for the subjective assessment of the quality of television pictures. 2012: International Telecommunication Union. Pág. 12-25
7. PÉCHARD, S., R. PÉPION, AND P. LE CALLET. Suitable methodology in subjective video quality assessment: a resolution dependent paradigm. In International Workshop on Image Media Quality and its Applications, IMQA2008. 2008. Pág. 6
8. BELLARD, F. AND M. NIEDERMAYER, FFmpeg. <http://www.ffmpeg.org>, 2012.