JITEL 2017
XIII Jornadas de Ingeniería Telemática
27-29 Septiembre 2017, Valencia (España)

*Actas de las XIII Jornadas de Ingeniería Telemática (JITEL 2017), Valencia (España), 27-29 de Septiembre de 2017*

ISBN: 978-84-9048-595-8
DOI: http://dx.doi.org/10.4995/JITEL2017.2017.6612

# Evaluación de un sistema DASH para el streaming de vídeo 3D

Paola Guzmán Castillo, Pau Arce Vila, Juan Carlos Guerri.
Grupo Comunicaciones Multimedia, iTEAM (Instituto de Telecomunicaciones y Aplicaciones Multimedia)
Universitat Politècnica de València
Camino de Vera, s/n.
paoguzc1,paarvi @iteam.upv.es, jcguerri@dcom.upv.es

***Resumen*- La distribución de contenidos multimedia, y en particular el streaming de vídeo, domina actualmente el tráfico global de Internet y su importancia será incluso mayor en el futuro. Miles de títulos se agregan mensualmente a los principales proveedores de servicios, como Netflix, YouTube y Amazon. Y de la mano del consumo de contenidos de alta definición que se convierte en la principal tendencia, se puede observar nuevamente un incremento en el consumo de contenidos 3D. Esto ha hecho que las temáticas relacionadas con la producción de contenidos, codificación, transmisión, calidad de servicio (QoS) y calidad de experiencia (QoE) percibidas por los usuarios de los sistemas de distribución de vídeo 3D sean un tema de investigación con numerosas contribuciones en los últimos años. Es importante tener en cuenta que en un sistema de distribución de vídeo las degradaciones debidas a la producción y la codificación, así como los errores de transmisión, puede degradar la calidad del vídeo recibida y percibida por el usuario. Por tanto, como parte de este trabajo se ha realizado en primer lugar una comparación del rendimiento de los estándares de codificación de vídeo más populares H.264, H.265 y sus correspondientes extensiones para vídeo 3D. Por otra parte, se ha realizado una evaluación experimental de la calidad del vídeo recibida en un escenario HTTP de streaming adaptativo (DASH) de vídeo 3D.**

***Palabras Clave*- DASH, vídeo 3D, HEVC, AVC, streaming adaptativo, QoE, QoS.**

## I. INTRODUCCIÓN

El servicio de streaming de vídeo crece y evoluciona a una velocidad increíble. Según las previsiones y estadísticas disponibles, el tráfico de vídeo representará el 82% de todo el tráfico de Internet para 2021, frente al 73% en 2016[1]. Por su parte, las recientes mejoras en la tecnología de vídeo 3D han suscitado nuevamente un creciente interés hacia el consumo de dichos contenidos, como una alternativa para expandir la experiencia del usuario. Nuevos contenidos demandan nuevos esquemas de representación y codificación, que se ajusten a las condiciones de transporte y restricciones de ancho de banda, y permitan maximizar la calidad de servicio (QoS, Quality of Service) y la calidad de experiencia (QoE, Quality of Experience) del usuario. En este sentido, tanto las perdidas asociadas a los procesos de codificación y compresión, como los errores y pérdidas durante la transmisión pueden afectar a la calidad percibida por el usuario.

Para poder estudiar el impacto que tiene cada uno de estos aspectos, nuestro primer objetivo ha sido realizar una comparación en términos de evaluación objetiva usando los parámetros PSNR (Peak Signal-to-Noise Ratio) y SSIM (Structural Similarity), empleando los estándares más populares de codificación de vídeo, como H.264/AVC (Advanced Video Coding) y H.265/HEVC (High Efficiency Video Coding) con sus respectivas extensiones para formatos multivista utilizando el estándar MVC (Multiview Video Coding).

En diversos estudios previos se han hecho comparativas de varias generaciones de estándares de codificación en términos de PSNR y pruebas subjetivas, e incluso comparando los nuevos codificadores de alta eficiencia H.265 y VP9. En un escenario de distribución de vídeo 3D, se utilizaban formatos estereoscópicos compatibles con 2D (Frame-compatible o Full-resolution Frame-compatible), pero la aparición de sistemas basados en pantallas auto-estereoscópicas o aplicaciones de telepresencia inmersiva ha motivado el desarrollo de formatos multivista como MVC o MVD (Multiview Video Plus Depth).

Otros factores que pueden afectar a la calidad de vídeo son las pérdidas y retardos ocasionados durante la transmisión. En la actualidad, el transporte de flujos de

vídeo en Internet se realiza cada vez más utilizando HTTP adaptativo, del cual el estándar DASH (Dynamic Adaptive Streaming over HTTP) es el protocolo más representativo.

Así, en [2] se evalúa el impacto del protocolo de transporte en la calidad percibida por el usuario cuando el vídeo se transmite sobre un enlace de ancho de banda limitado, utilizando YouTube como ejemplo. Por otro lado, en [3] se presenta un análisis comparativo de las diferentes estrategias de adaptación de bitrate en un escenario de streaming adaptativo tanto monoscópico como estereoscópico.

Asimismo, en publicaciones recientes [4] se revisan de forma exhaustiva los métodos de evaluación tanto objetivos como subjetivos relacionados con el servicio de videostreaming. Por su parte en [5] se aborda la evaluación de la QoE en escenarios de vídeo 3D comparando diferentes técnicas de codificación y utilizando como referencia la recomendación BT.500.13 ITU-R.

En este sentido, se plantea evaluar el efecto que produce la adaptación de la tasa de vídeo 3D al ancho de banda disponible y cuál es su influencia en la QoE, utilizando el escenario de la Fig. 1.

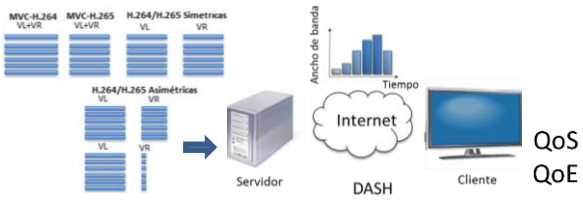



Fig. 1. Diagrama general.

Por tanto, en la sección II se hará una descripción del procedimiento seguido para la selección de secuencias y los resultados de la comparación de los estándares H.264/AVC y H.265/HEVC. En la Sección III se mostrarán los resultados de la evaluación experimental de un sistema de DASH para vídeo 3D. Finalmente, las conclusiones se exponen en la Sección IV.

## II. Selección de Sequencias y Comparación de Codificadores

### A. Selección de secuencias

Las secuencias utilizadas en este trabajo han sido tomadas de las bases de datos de vídeos 3D HD estereoscópico, Nantes-Madrid-3D-Stereoscopic-V1, NAMA3DS1 y RMIT3DV, que se encuentran disponibles abiertamente en sus sitios web. Dichas bases de datos están compuestas por secuencias estereoscópicas (vistas por separado) con resolución 1920x1080 a 25 fps, diseñadas para representar una amplia gama de contenidos y condiciones visuales. Asimismo, se ha empleado la ya popular secuencia de animación Big Buck Bunny en 3D producida por Blender Foundation. Se evaluarán un total de 13 secuencias con duración entre 16 s y 634 s la más larga.

Con el fin de cuantificar y evaluar la variedad de los contenidos seleccionados y la dificultad de codificación, tal como se describe en la Recomendación ITU-T P.910, se ha calculado el índice de información espacial (SI) y el índice de información temporal (TI) de cada una da las secuencias bajo estudio. En la Fig. 2 se muestran los índices SI y TI calculados sobre la componente de luminancia de cada secuencia y una miniatura de cada una de ellas.

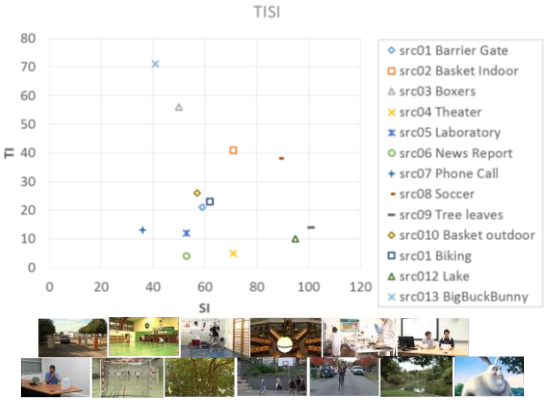



Fig. 2. Indices SI –TI

Dados los tiempos de codificación de algunas de las herramientas que serán empleadas para realizar la comparación entre los estándares H.264/AVC y H.265/HEVC, para esta primera fase, se seleccionaron 10 s de cuatro de las secuencias mencionadas anteriormente. Intentando cubrir un espectro amplio en cuanto a tipo de contenidos, se ha elegido una de tipo outdoor, una indoor, una de deportes y otra de animación.

### B. Comparación de codificadores

Para la comparación de los estándares H.264/AVC y H.265/HEVC en el ámbito de la codificación de secuencias estereoscópicas, además del software libre ffmpeg-libx264 y ffmpeg-libx265, se emplearon los codificadores de referencia. Por un lado, HEVC test Model (HM 16.7) para las codificaciones HEVC Simulcast y MVC-HEVC, y por otro lado, los codificadores JM19.0 y JMVC 8.5 para las codificaciones AVC simulcast y MVC-AVC, respectivamente.

Para generar variaciones de calidad, se emplearon los parámetros de cuantificación (QP) 24, 28, 32, 36 y 40, pero debido a la diferencia significativa en las capacidades de los codificadores, se buscó que los parámetros de configuración fuesen equivalentes en todos los codificadores, eligiendo los mismos valores de GoP (Group of Pictures) y separación entre frames Intra.

La Fig. 3 muestra las curvas RD (Rate-Distortion) en función del PSNR, para dos de las secuencias bajo estudio. Para ésta y las figuras siguientes usaremos PSNR como medida de calidad. PSNR es la métrica más utilizada en la compresión de vídeo ya que, aunque no siempre refleja la calidad perceptual, es una manera simple de medir la fidelidad a la fuente (40 dB o

superior es muy buena calidad, por debajo de 35 dB mostrará artefactos de codificación).


Fig. 3. Curva RD (PSNR) comparación de codificadores.

Como se puede ver en la Fig. 3, en todos los casos evaluados los codificadores basados en HEVC presentan una mejor relación entre la calidad del vídeo obtenida y el bitrate. También podemos observar que los codificadores de referencia, en particular las versiones MVC, tienen mejores prestaciones comparados con los resultados obtenidos al codificar cada vista por separado. Esto se debe a que además de considerar las similitudes entre frames dentro de cada vista, explotan también la similitud intervista, lo que permite suprimir frames de tipo Intra al codificar una de las vistas en función de la otra.

Sin embargo, enfocados en nuestro siguiente objetivo que se centra en la transmisión de contenidos de vídeo 3D en un entorno adaptativo, además de la eficiencia de codificación, se ha valorado la opción de poder generar codificaciones asimétricas y los tiempos de codificación. La Tabla II muestra la amplia diferencia entre los tiempos de codificación usando codificadores de referencia respecto al ffmpeg. Las codificaciones asimétricas se valen de las características del sistema de visión humano, el cual al visualizar un vídeo 3D donde cada una de las vistas ha sido codificada con un factor de calidad, se ha demostrado que prevalece la sensación de la vista con mejor calidad [6] y resultan una buena alternativa para la reducción de la tasa de bits en función de la calidad en un escenario de streaming adaptativo.

Tabla II
TIEMPOS DE CODIFICACIÓN- COMPARACIÓN DE CODIFICADORES DE REFERENCIA Y FFMPEG (MVC VS VISTAS POR SEPARADO)

| Codificador | Tiempo de codificación (s) | Segundos por frame |
|---|---|---|
| *FFMPEG H.264* | 121 | 0,4 |
| *JM H.264* | 23743 | 79,1 |
| *JMVC MVC-H.264* | 23778 | 79,3 |
| *FFMPEG H.265* | 3179 | 10,6 |
| *HM H.265* | 31382 | 104,6 |
| *HM MVC-H.265* | 1595 | 5,3 |

Por tal motivo, en adelante se usará como herramienta para la codificación ffmpeg con las librerías de codificación libx264 y libx265. La Tabla II muestra el promedio del tiempo de codificación con cada aplicación. El ordenador donde se realizaron las codificaciones es un Pentiun Dual-Core a 2.5GHz con una versión Ubuntu 16.10 de 64 bits.

De los resultados obtenidos en un trabajo anterior [6], donde se emplearon métodos de evaluación tanto objetivos como subjetivos de calidad de vídeo, se pudo demostrar que el estándar H.265/HEVC proporciona un ahorro significativo en el bitrate con respecto a H.264/AVC. En línea con estos resultados, para las secuencias tomadas en el presente trabajo, tal como se muestra en la Tabla III, la reducción en la tasa de bits para una misma calidad varía entre un 66,8% a un factor de calidad bajo (QP40) para una secuencia de tipo indoor con bajo nivel de movimiento (src06 – informativo noticias), y un 10,9% para un factor de calidad alto QP24 y una secuencia de deporte (src010 – Basket outdoor). En la secuencia de tipo animación (src013 – BigBuckBunny) la reducción de bitrate presenta una variación menor, oscilando entre un 36,9% con QP24 y un 59,3% con QP40.

Tabla III
REDUCCIÓN EN LA TASA DE BITS DE H.265/HEVC (FFMPEG) RESPECTO A H.264/AVC (FFMPEG)

| Secuencia | QP24 | QP28 | QP32 | QP36 | QP40 |
|---|---|---|---|---|---|
| *Src01* | 23,4% | 42,8% | 48,8% | 52,6% | 55,7% |
| *Src06* | 36,0% | 56,2% | 62,6% | 66,0% | 66,8% |
| *Src010* | 10,9% | 36,2% | 41,7% | 45,0% | 49,0% |
| *Src013* | 36,9% | 48,1% | 53,2% | 56,1% | 59,3% |

## III. EVALUACIÓN DE LA TRANSMISIÓN CON DASH

### *A. Producción de contenidos DASH*

Como paso previo a la generación de los contenidos DASH se debe realizar un proceso de codificación, que como se mencionó anteriormente en nuestro caso se ha realizado empleando la aplicación ffmpeg y las librerías libx264 y libx265. Como parámetro de codificación se ha empleado la tasa de bits máxima. Como paso previo a la generación de las secuencias en formato SBS (Side by Side) Frame-compatible, las secuencias correspondientes a cada una de las vistas (izquierda y derecha) son codificadas con diferentes valores de bitrate (2000kbps, 1600kbp, 1000kbps, 700kbps, 500kbps). Teniendo en cuenta que no todas las secuencias se comportan igual frente a determinados parámetros de codificación, la selección de la tasa de bits óptima es un tema a tener en cuenta. Al realizar un gráfico comparativo que incluye las curvas RD de las 13 secuencias codificadas con 5 valores diferentes de QP, se observa como mientras algunas secuencias de vídeo alcanzan un PSNR muy alto (45 dB o más) a bitrates de 2000 kbps o menos, por otra parte, algunos vídeos a estas tasas de bits sólo alcanzan un valor aceptable de PSNR de 38 dB.

Cuando se utiliza el estándar DASH, el vídeo se segmenta de forma que el cliente pueda pedir los

segmentos uno a uno, y en función del ancho de banda que mida en cada momento, pueda escoger descargar los siguientes segmentos de vídeo de mayor o menor calidad. De entre las secuencias evaluadas en la Sección II, se han seleccionado las codificaciones óptimas a partir de la curva de PSNR, maximizando la relación de PSNR y bitrate.

Una vez codificadas las secuencias, se utiliza MP4Box para convertir el vídeo en segmentos DASH de 5 segundos y generar un archivo de índice MPD (Media Presentation Description), que contiene toda la información sobre las diferentes calidades de vídeo usadas y los anchos de banda de cada una. Este es el archivo que el cliente utilizará para saber qué segmentos descargarse en función del ancho de banda medido. Finalmente, como servidor de contenidos hemos usado un equipo Apache 2.4.18 para Linux Ubuntu. De entre todos los clientes disponibles para reproducir DASH, se utiliza una versión modificada por los autores del Shaka Player V2.0.0, del que se puede obtener información relativa al throughput, tiempo de descarga y nivel del buffer de reproducción.

### *B. Evaluación prestaciones DASH*

Para evaluar el comportamiento de adaptación de DASH y cómo afecta a la calidad de vídeo 3D, se emulan diferentes canales de transmisión con ancho de banda variable. Para ello se utiliza la herramienta NetEm, que es capaz de modificar y restringir el ancho de banda de salida del servidor (también el retardo o la pérdida de paquetes) de forma que el cliente perciba cambios en la tasa de descarga instantánea y adapte la calidad de vídeo en consecuencia.

Para los experimentos se han definido diferentes canales de transmisión, variaciones rápidas y lentas de ancho de banda, teniendo en cuenta el tamaño de segmento utilizado (5 s). Por cuestiones de espacio, en este trabajo se presenta únicamente el escenario con variaciones de ancho de banda cada 40 s.

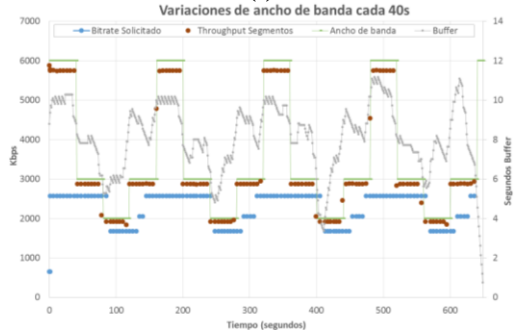
La Fig. 4(a) muestra el throughput por segmento alcanzado en un escenario sin restricciones en el que el cliente puede descargar las representaciones de más alta calidad. El archivo MPD ofrece 9 calidades entre 0,65 Mbps y 2,5 Mbps. Por su parte en la Fig. 4(b) se muestra el comportamiento del Shaka Player frente a un escenario de variaciones persistentes de ancho de banda.

## IV. CONCLUSIONES

El uso de codificaciones asimétricas para la representación de las secuencias de vídeo estereoscópico resulta de especial interés en un escenario HTTP de streaming adaptativo (DASH) ya que, aprovechando las características del sistema audiovisual humano donde prevalece la sensación de la vista con mejor calidad, permite reducir el ancho de banda de las transmisiones y aumentar la granularidad en los cambios de calidad, mejorando así la calidad subjetiva percibida por el usuario.

El mecanismo de adaptación de Shaka Player, además de evitar el congelamiento del vídeo frente a una caída del ancho de banda, hace un buen uso del buffer en situaciones de reducción de ancho de banda, lo cual le permite minimizar el número de cambios de calidad efectuados en un escenario de variaciones de ancho de banda con alta frecuencia, al mismo tiempo que reacciona con gran rapidez en situaciones de aumento de ancho de banda.


(a)




(b)

Fig. 4. (a) Throughput y el buffer por segmentos sin restricciones de ancho de banda. (b) Ancho de banda, Throughput por segmento, bitrate solicitado, ocupación de buffer de reproducción

## AGRADECIMIENTOS

Este artículo se enmarca en el Proyecto PROMETEOII/2014/003 financiado por la Generalitat Valenciana y el Proyecto "Desarrollo de Nueva Plataforma de Entretenimiento Multimedia para Entornos Náuticos" (CDTI IDI -20170348).

## REFERENCIAS

[1] Cisco, "Cisco Visual Networking Index: Forecast and Methodology, 2016-2021,", 2017.

[2] T. Hoßfeld, R. Schatz, and U. R. Krieger, "QoE of YouTube Video Streaming for Current Internet Transport Protocols," in *Measurement, Modelling, and Eval. of Computing Systems and Dependability and Fault Tolerance*, 2014, pp. 136–150.

[3] S. Tavakoli, J. Gutierrez, and N. Garcia, "Subjective quality study of adaptive streaming of monoscopic and stereoscopic video," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 4, pp. 684–692, 2014.

[4] Y. Chen, K. Wu, and Q. Zhang, "From QoS to QoE: A Tutorial on Video Quality Assessment," *IEEE Commun. Surv. Tutorials*, vol. 17, no. 2, pp. 1126–1165, 2015.

[5] T. Tian, X. Jiang, and X. Du, "Subjective quality assessment of compressed 3D video," in *2014 7th International Congress on Image and Signal Processing*, Dalian, 2014,

pp. 606–611.

[6] P. Arce, I. De Fez, F. Fraile, S. González, P. Guzmán, and J. C. Guerri, "QoE en redes adhoc, descarga adaptativa de contenidos y vídeo 3D,"*Proc. of Jornadas de Ingeniería Telemática (JITEL)*, Mallorca, 2015.