



Codificación de video en HEVC/H.265 utilizando FFMPEG

Rafael Alejandro Olivera Solís, Yazdani López Pérez

RESUMEN / ABSTRACT

HEVC/H.265 es un nuevo estándar de codificación de video. Su objetivo principal es mejorar los niveles de compresión de video. La compresión de video ha jugado un papel fundamental en el desarrollo de las aplicaciones multimedia. La necesidad de reducir el bitrate, y a su vez mantener la calidad lo mayor posible, ha traído consigo la utilización de herramientas que permitan un proceso de codificación eficiente y eficaz. Este trabajo realiza una estimación de los parámetros relativos a la compresión eficiente de señales de video, así como las métricas de calidad objetiva; a partir de la modificación de los parámetros de codificación de la señal, utilizando potentes herramientas de software libre y un clúster de alto rendimiento.

Palabras claves: Compresión, FFMPEG, HEVC/H.265, HPC

HEVC / H.265 is a new standard for video encoding. Its main objective is to improve video compression levels. Video compression has played a fundamental role in the development of multimedia applications. The need to reduce the bit rate, and in turn maintain the highest possible quality, has brought the use of tools that enable efficient and effective process of coding. This work makes an estimation of the parameters related to the efficient compression of video signals, as well as objective quality metrics; from the modification of the coding parameters of the signal, using powerful free software tools and a high performance cluster.

Keywords: *Compression, FFMPEG, HEVC/H.265, HPC*

Estimation of objective video quality using HEVC / H.265.

1. -INTRODUCCIÓN

En la sociedad actual es innegable la importancia que tiene el video como soporte de información, tanto en ámbitos empresariales (videoconferencias y seguridad) como de ocio (cine, televisión, entre otros). Redes Sociales como Facebook, MySpace y servicios como YouTube realizan cada vez un mayor uso del video en su oferta.

La difusión por satélite, la Televisión por Protocolo de Internet (IPTV) y la Televisión Digital Terrestre (TDT), son sistemas donde los recursos del canal de comunicaciones son limitados. Por tanto, es necesario emplear técnicas de compresión que reduzcan de forma drástica el régimen binario de la señal de video codificada, manteniendo la misma calidad visual. En 2013 se estandariza HEVC/H.265 (*High Efficiency Video Coding*), un formato de compresión de video de alta eficiencia. Este estándar fue desarrollado por ITU-TVCEG (*Video Coding Experts Group*) e ISO/IEC MPEG (*Moving Picture Experts Group*). Su objetivo principal fue mejorar la compresión de video, en relación a los anteriores estándares [1,2].

El estándar HEVC permite un importante avance en la compresión con respecto a sus predecesores [3]. Las aplicaciones para HEVC no solo cubren el espacio de los conocidos usos actuales y capacidades del video digital, sino que también incluyen el despliegue de nuevos servicios, la entrega de una calidad de video mejorada y de imágenes de Televisión de Ultra Alta Definición (UHDTV) con video de mayor rango dinámico, una gama más amplia de color representable y mayor

precisión de representación [3]. El manejo del estándar HEVC/H.265 es una tarea engorrosa para ordenadores simples, incluso con buenas prestaciones. Por este motivo, el empleo de un clúster de alto rendimiento (HPC), en el que se distribuya el procesamiento entre todos los procesadores de un mismo nodo, constituye una solución ante la problemática del procesamiento de la señal de video en el proceso de codificación, siempre que los requerimientos de inmediatez lo permitan. Dentro de este, la utilización de la plataforma FFMPEG es otra de las soluciones poco exploradas en el manejo de este estándar. Esta herramienta fue la primera que contó con la librería x265, además de brindar múltiples bondades para el trabajo con la señal de video. A lo anterior se le adiciona que FFMPEG permite obtener en el propio proceso de codificación las métricas de calidad objetiva, ahorrando recursos computacionales y tiempo de procesamiento; en contraposición con aplicaciones comerciales, donde es necesario poseer la señal codificada y la muestra sin codificar. En estos casos los tiempos para obtener resultados concretos aumentan considerablemente. Todo ello sin tener en cuenta que las aplicaciones más reconocidas no son libres de pago.

El auge de teléfonos inteligentes, computadoras portátiles o tabletas y otros dispositivos que incorporan funciones para la captura de imágenes y videos han inundado, de contenido multimedia de alta resolución, la Web. Dicho contenido presenta características variables en cuanto a resolución, número de fotogramas por segundo, relación de aspecto, tasa de bit y formato contenedor multimedia de video. En muchas ocasiones, los inconvenientes con la transmisión, difusión, almacenamiento y/o reproducción del video digital de alta resolución encuentran solución en la compresión eficiente de datos. Pero no siempre la compresión eficiente se puede realizar debido a que los estándares de codificación imponen limitaciones en la precisión de las operaciones de muchos de los procesos internos del codificador, lo que provoca errores de redondeo que a su vez desembocan en una pérdida de eficiencia de codificación perceptible cuando se evalúan las medidas objetivas de calidad visual PSNR (*Peak Signal-to-Noise Ratio*) y SSIM (*Structural Similarity*).

El artículo realiza aportes en la comprensión del proceso de codificación de la señal de video. Realiza la estimación de los parámetros relativos a la compresión eficiente de video, así como las métricas de calidad PSNR y SSIM, tomando como base la variación de los parámetros de codificación del estándar HEVC/H.265 sobre la plataforma FFMPEG dentro de un clúster de alto rendimiento. Se realiza un análisis exhaustivo de los resultados obtenidos tomando como base algunas de las herramientas que están presentes en los codificadores x.265 de la plataforma FFMPEG. El trabajo se organiza de la siguiente manera. A continuación, se caracteriza el estándar HEVC/H.265, haciendo énfasis en los nuevos elementos que contiene. Luego se definen las métricas de compresión eficiente y se caracteriza el entorno de codificación tanto desde el punto de vista de hardware como de software. Se describen los experimentos realizados y se muestran los resultados obtenidos. Al final, en las conclusiones, se realiza un análisis crítico de los resultados obtenidos en el trabajo.

2.- MATERIALES Y MÉTODOS

2.1.- HEVC/H.265

Como parte de los avances en el campo de la codificación de fuente llevados a cabo por la comunidad científica, se aprueba la primera fase del estándar de compresión H.265/HEVC a inicios del 2013. Desde ese momento se estimaba que fuera el sucesor natural del estándar de compresión H.264/AVC (*Advanced Video Coding*). En la figura 1 se muestra el diagrama en bloques de un codificador-decodificador HEVC/H.265.

HEVC/H. 265 mejora las herramientas existentes en los estándares previos, específicamente en H.264, logrando una mejora en la calidad percibida y permitiendo una reducción de bitrate por encima del 50% con respecto a H.264. H.265 pretende ser altamente eficiente en dos sentidos: lograr eficiencia en formatos de muy alta resolución como UHD TV (*Ultra High Definition Television*), y hacerlo también en entornos de muy baja tasa binaria, como es el caso de servicios *streaming*, WebTV y OTT (*Over the Top Content*). H.265, además, utiliza la estimación y la compensación de movimiento para aprovechar la similitud temporal de las secuencias, y además emplea la transformada discreta del coseno, con cierta variación, con el objetivo de aplicar la cuantificación en el dominio transformado, y codificar estadísticamente dichos coeficientes. Para ello H.265 cuenta con diferentes estructuras novedosas que se analizarán a continuación.

- Nueva Unidad de Codificación: Define una nueva estructura de codificación que diverge de las anteriores unidades denominadas macro bloques con dimensiones de 16x16 píxeles. En este estándar la unidad de codificación se denomina *Coding Tree Block* (CTB) y sus dimensiones pueden ser desde 8x8 hasta 64x64 píxeles. En HEVC, una imagen se

divide en diferentes CTU (*Coding Tree Units*) que pueden presentar un tamaño máximo de $L \times L$ píxeles, para el caso de la luminancia L puede tomar los valores de 16, 32 ó 64, con los tamaños más grandes que permiten típicamente una mejor compresión [5].

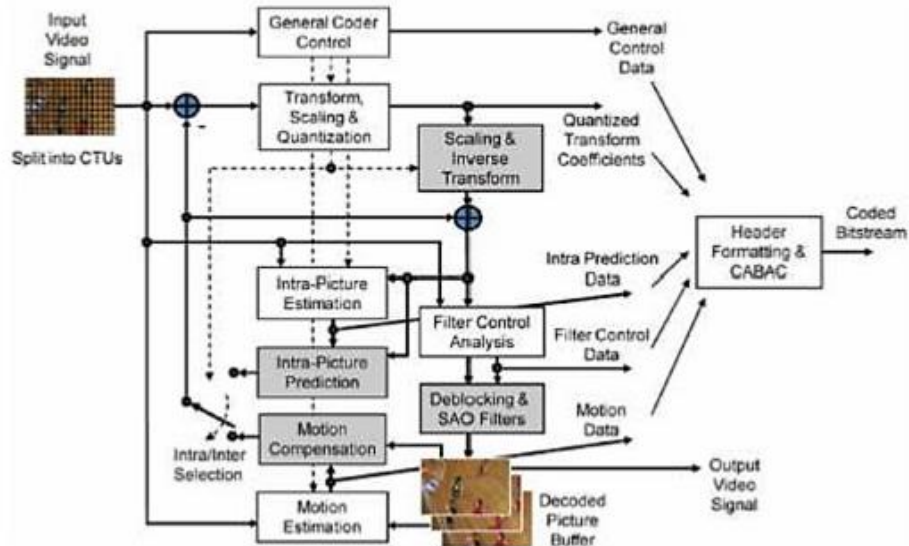


Figura 1

Diagrama en bloques del codificador HEVC/H.265 [4].

- Unidad de Predicción: H.265 define una nueva unidad de predicción (PU) que puede tener dimensiones desde 32×32 hasta 4×4 , lo que representa la mitad que la menor de las unidades de codificación. Al igual que su antecesor H.264, pueden aplicarse de modo intra-frame o inter-frame, utilizando divisiones distintas.
- Predicción Intra-Frame: HEVC introduce un nuevo método de predicción intra-frame. El método intra-angular define 33 predicciones direccionales para todos sus posibles tamaños de PU el cual se diferencia de los 2 predictores direccionales que define H.264 para tamaños de 16×16 , y 8 predictores para 8×8 y 4×4 . Además, se definen los modos intra-planar, donde se supone una superficie de amplitud con una pendiente horizontal y vertical derivada de los límites, y el modo intra-DC donde se asume una superficie plana con un valor que coincida con el valor medio de las muestras de contorno [6].
- Predicción Inter-Frame: La estimación de movimiento en H.265 se lleva a cabo con la precisión de $\frac{1}{4}$ de píxel con filtros interpoladores de 7-taps y 8-taps. Se incorpora además la AMVP (*Advanced Motion Vector Prediction*) para lograr la señalización de sus vectores de movimiento mediante el cálculo de los vectores de movimiento más probables obtenidos desde los bloques vecinos [7].
- Unidades de Transformación: Se define una nueva TU (*Transform Unit*). Se aplica de igual forma la DCT (*Discrete Cosine Transforme*) con una ligera modificación con respecto a la utilizada en H.264, también permite bajo ciertas condiciones la utilización de la DST (*Discrete Sine Transforme*). En el caso de H.265 se amplían los tamaños de transformación desde 4×4 a 32×32 .
- *Sample Adaptive Offset* (Compensación Adaptable de la muestra): Nueva herramienta aplicada luego del *In-loop filter* con el objetivo de lograr mejorar la calidad subjetiva de la imagen decodificada, aplicando pequeños *offset* a los píxeles decodificados, en función de la zona a la que pertenecen [1].

Debido a la alta complejidad de los algoritmos introducidos por H.265, el estándar integra novedosas herramientas con el objetivo de viabilizar la codificación de la imagen. HEVC permite la codificación del Frame en dos nuevos tipos adicionales al *slice*, que existía previamente en los anteriores estándares, denominados *Tile* y Frente de Ondas o WPP (*Wavefront Parallel Processing*).

- *Tile*: define regiones descifrables rectangulares de la imagen, típicamente con el mismo número de CTU, éstas pueden ser codificadas independientemente y comparten algunas cabeceras de información. Los *tiles* tienen como propósito fundamental aumentar la capacidad para el procesamiento paralelo en lugar de proporcionar la capacidad de recuperación de errores. Múltiples *tiles* pueden compartir información de cabecera por estar contenida en el mismo *slice*. Alternativamente, un solo *tile* puede contener múltiples *slices*. Los *tiles* proporcionan cierto nivel de paralelización en los procesos de codificación y decodificación de video que puede ser explotado en sistemas con arquitecturas de procesamiento paralelo, siendo más flexibles que las *slices* de H.264. Además, se considera una tecnología menos compleja que FMO (*Flexible Macroblock Ordering*) [8].
- WPP: Es una herramienta utilizada por HEVC en el procesamiento en paralelo. Cuando WPP (*Wavefront Parallel Processing*) se encuentra habilitado, una *slice* se fracciona en filas de CTU. La primera fila puede ser procesada por un hilo de ejecución; en el caso de un segundo hilo de ejecución puede comenzar con la segunda fila después de haberse procesado las dos primeras CTU de la fila anterior. Al igual que los *tiles*, las WPP suministran una forma de paralelización en los procesos de codificación y decodificación. La diferencia radica en que con WPP se consigue un mejor rendimiento de compresión además de no permitir la introducción de determinadas aberraciones visuales como ocurre en los *tiles* [9].

2.2.- MÉTRICAS DE CALIDAD OBJETIVAS

Las primeras medidas objetivas de la calidad del video utilizadas son del tipo FR (*Full Reference*) y están basadas en obtener las diferencias, píxel a píxel, entre las imágenes originales (previo a la compresión y transmisión) y las imágenes degradadas (luego de la recepción y reconstrucción). Las medidas más simples de este tipo son las de error cuadrático medio (MSE - *Mean Square Error*) y el valor pico de la relación señal a ruido (PSNR *Peak Signal to Noise Ratio*), definidas en las ecuaciones (1) y (2). Como se mencionó, estas métricas son del tipo FR, ya que requieren de la referencia completa para poder ser calculadas.

El PSNR es una de las medidas de calidad visual objetiva más utilizadas desde los comienzos de la codificación de video digital. Define la relación entre la máxima energía posible de una señal y el ruido que afecta a su representación fidedigna. Dado el gran rango dinámico de las señales de video, donde el valor de cada muestra puede ir desde 0 hasta 255 para videos de 8 bits por muestra y desde 0 hasta 1023 para videos de 10 bits por muestra, su valor se suele representar en escala logarítmica, con el decibel (dB) como su unidad. El PSNR se calcula mediante la siguiente expresión [10]:

$$PSNR = 10 \log \left(\frac{MAX^2}{MSE} \right) \quad (1)$$

donde el MSE de las dos imágenes a comparar se calcula mediante la expresión:

$$MSE = \frac{1}{MN} \sum_{I=0}^{M-1} \sum_{J=0}^{N-1} |O(i, j) - D(i, j)|^2 \quad (2)$$

en la cual M y N son el ancho y el alto de la imagen en píxeles, respectivamente, y donde O (i, j) y D (i, j) son cada píxel situado en la fila i y la columna j de la imagen original y distorsionada, respectivamente.

MAX representa el valor máximo que puede tomar cada píxel. Este valor depende del número de bits por muestra de la señal, y como se trabaja con señales que poseen diferente número de bits por muestra, B se define como:

$$MAX = 2^B - 1 \quad (3)$$

lo cual equivale a 255 para señales de 8 bits por muestra, y 1023 para señales de 10 bits por muestra.

Los valores típicos de PSNR en una compresión con pérdidas de imagen y video oscilan entre 30 y 50 dB, para una profundidad de 8 bits, donde un mayor valor de PSNR es resultado de una mejor compresión. Para 16 bits de profundidad, el valor del PSNR oscila entre 60 y 80 dB.

Debido a la baja correlación entre el MSE y PSNR con la calidad percibida, en los últimos tiempos, se ha realizado un gran esfuerzo para desarrollar nuevos modelos que tengan en cuenta las características de percepción del sistema visual humano y que permitan calcular métricas objetivas que lo simulen.

El SSIM es una medida de calidad visual objetiva, creada en 2004, que surgió con la finalidad de contar con un parámetro de calidad objetivo más cercano al sistema visual humano.

La manera más sencilla de entender esta medida es la siguiente: mientras que el PSNR es una medida basada en el error entre dos señales, SSIM es una medida de distorsión estructural. Para su desarrollo el autor aprovechó el hecho de que una de las funciones principales del ojo humano es extraer información estructural del campo de visión, ya que los humanos tienen una tendencia muy desarrollada a visualizar formas y contornos de manera precisa. SSIM considera la degradación de la imagen como pérdida de la información estructural percibida.

El cálculo de SSIM de una imagen se realiza en ventanas de pequeño tamaño, normalmente de 8x8 píxeles mediante la siguiente expresión [10,11]:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c1)(2\sigma_{xy} + c2)}{(\mu_x^2\mu_y^2 + c1)(\sigma_x^2\sigma_y^2 + c2)} \quad (4)$$

$$c1 = (0,01 \times MAX)^2 \quad (5)$$

$$c2 = (0,03 \times MAX)^2 \quad (6)$$

siendo MAX el valor máximo que puede tomar cada píxel, 255 para señales de 8 bits por muestra y 1023 para señales de 10 bits por muestra. Una vez realizado el cálculo de n ventanas, cuyo número dependerá de la resolución de la imagen y del algoritmo utilizado para el cálculo, el valor final de SSIM es la media aritmética de los valores de todas las ventanas.

Los valores de SSIM pueden ir desde 0 hasta 1. Cuanto más se acerque el valor a 1 menos distorsión estructural habrá entre las dos imágenes, siendo el valor 1 solo obtenible en el caso de dos imágenes exactamente iguales. SSIM suele usarse solo con los planos de luminancia de cada fotograma, debido a la poca información estructural contenida en los planos de diferencia de color [11].

3.- RESULTADOS Y DISCUSIÓN

El experimento realizado se basa en el procesamiento de tres secuencias de video sin compresión de diferentes resoluciones. En la tabla # 1 se exponen las características de las muestras de video.

Tabla 1
Características de las muestras de video sin compresión

| Parámetros/Muestras | Resolución Estándar | Resolución FullHD | Resolución 4K |
|----------------------|---------------------|-------------------|----------------|
| PROFUNDIDAD DE BITS | 8 bits | 8 bits | 10 bits |
| BITRATE | 291 964 Kbps | 622 080 Kbps | 4 150 263 Kbps |
| FORMATO CONTENEDOR | YUV4MPEG2 | YUV4MPEG2 | YUV4MPEG2 |
| RELACIÓN DE ASPECTO | 4:3 | 16:9 | 1.896:1 |
| FRAME RATE | 60 fps | 25 fps | 60 fps |
| CANTIDAD DE IMÁGENES | 600 | 217 | 600 |
| SUBMUESTREO DE COLOR | 4:2:0; S=1,5 | 4:2:0; S=1,5 | 4:2:0; S=1,5 |
| TAMAÑO | 348 MB | 643 MB | 14,8 GB |

| | | | |
|----------|------|-----|------|
| DURACIÓN | 10 s | 9 s | 10 s |
|----------|------|-----|------|

Esta tarea se realizó utilizando la plataforma FFMPEG de software libre en uno de los nodos del clúster de alto rendimiento de la Universidad Central “Marta Abreu” de las Villas. Las características de hardware del nodo utilizado se exponen a continuación:

- **Nodo:** IBM IDataPLex dx360 M2.
- **Procesadores:** 8 CPUs Intel Xeon L5520, (*Gainestown or Nehalem-EP*).
- **Frecuencia:** 2,394 GHz.
- **Caché:** 4 L2 de 265 Kb, L3 de 8 MB.
- **Memoria RAM:** 64,82 GB.

Los recursos que demandó realizar el proceso de codificación en el nodo del clúster se obtuvieron de los datos disponibles en el sitio <http://ganglia.uclv.hpc.cu/>. Estos corresponden al porcentaje de utilización de las CPUs en el nodo utilizado, la cantidad promedio de hilos de procesamiento (*Threads*) que demanda la tarea y la memoria RAM necesaria para acometer el proceso de codificación. En la figura 2 se muestra el comportamiento del nodo en cuanto al porcentaje de utilización de CPUs para la codificación de las 3 muestras incluyendo los 3 perfiles de trabajo en cada una de las codificaciones.

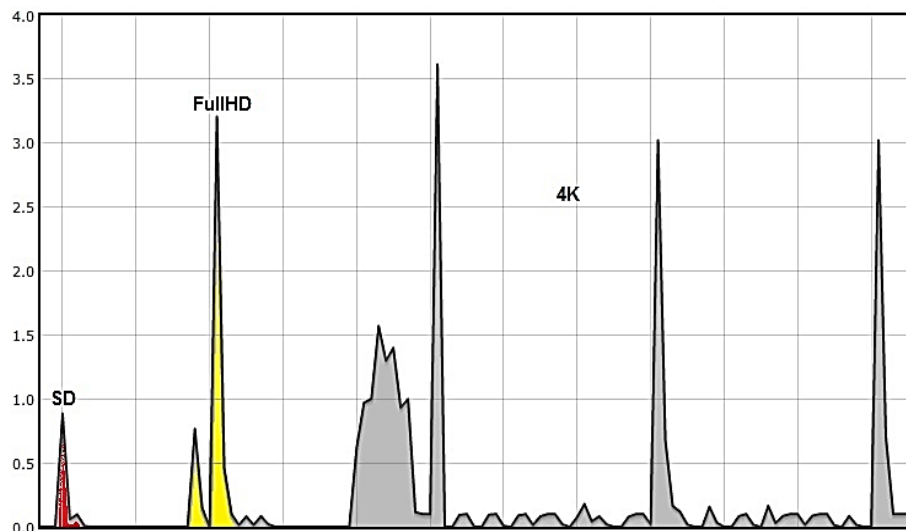


Figura 2

Porcentaje de uso de las CPUs del nodo del clúster.

En la figura 3 se muestra la cantidad de hilos (*Threads*) que se utilizaron como promedio para acometer el proceso de codificación de cada una de las muestras en cada uno de los 3 perfiles de trabajo.

En la figura 4 se puede observar el monto de memoria RAM que fue utilizada para realizar las codificaciones. Es necesario aclarar que la figura muestra la diferencia entre la memoria RAM en estado de inactividad, o la máxima memoria disponible en el nodo, y la disminución de la misma a partir de la ejecución de una tarea determinada.

Al analizar los gráficos de rendimiento del nodo del clúster se puede apreciar la diferencia en cuanto a requerimientos computacionales que demanda cada muestra de video. En cuanto a los perfiles de trabajo, en el caso de la muestra de resolución estándar es prácticamente imposible de notar el cambio de un perfil a otro respecto a la utilización de las CPUs, ya que las codificaciones ocurren en tiempos relativamente cortos. En el caso de la cantidad de hilos (*Threads*) utilizados como promedio para cada tarea, se evidencia que la muestra en resolución estándar solo utiliza entre 5 y 6 hilos. En cambio las otras muestras sí utilizan la totalidad de los hilos disponibles durante el proceso de codificación. La gran diferencia está en el monto de memoria RAM que demandan en el proceso. Para la muestra 4K fue necesaria la utilización de aproximadamente 20 GB de memoria RAM. La muestra FullHD necesitó aproximadamente 2,1 GB, mientras que la muestra de resolución estándar demandó del nodo 1,4 GB aproximadamente. De manera general, se evidencia la necesidad de acometer la tarea de codificación con H.265/HEVC en computadoras con altas prestaciones. Con las muestras empleadas, se pone de manifiesto que para el proceso de codificación lo primordial es la disponibilidad de memoria RAM, ya que, como se aprecia en la figura 2, el porcentaje de uso de las CPUs, no es elevado.

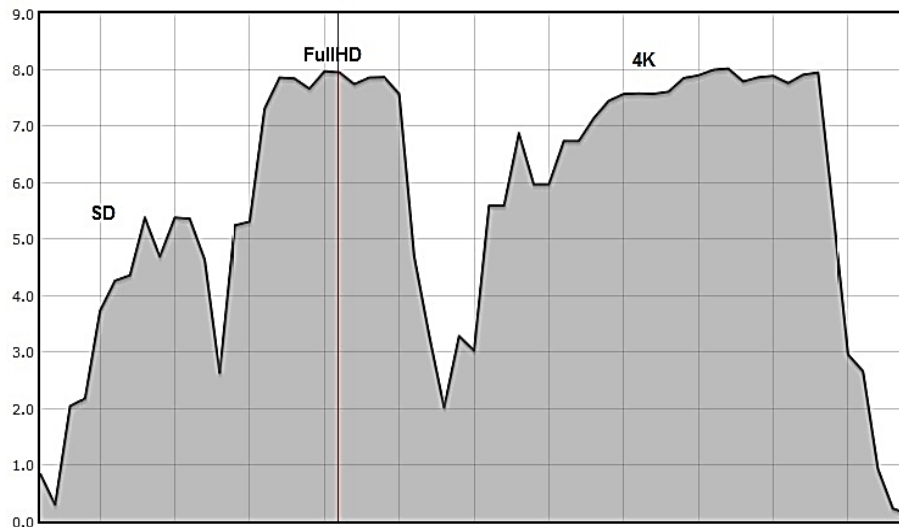


Figura 3

Cantidad promedio hilos de procesamiento (*Threads*) utilizados por muestra de video.

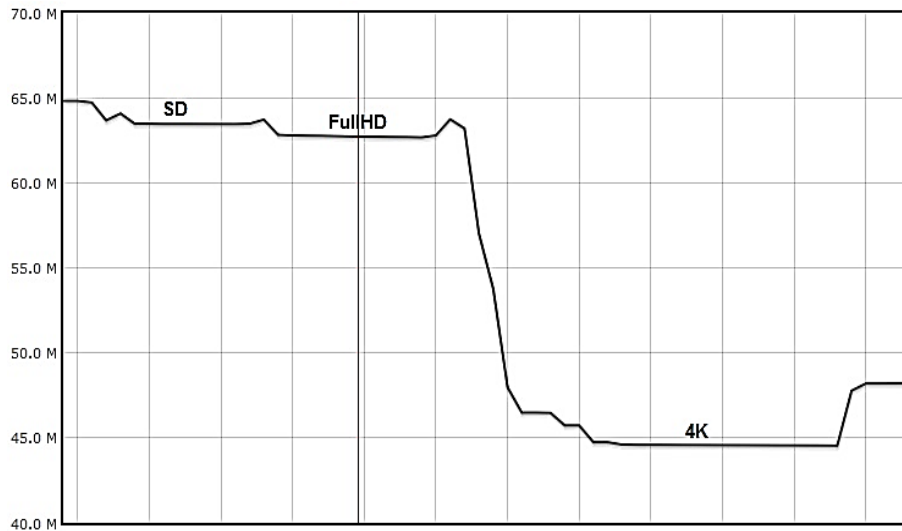


Figura 4

Memoria RAM utilizada en el proceso de codificación.

Las tres muestras poseen resoluciones diferentes y en este sentido, el códec H.265 analiza las posibilidades de que existan grandes zonas en la imagen con características similares. Esto se aprovecha en mayor medida en las resoluciones 1080p y 4K, donde se explotan los tipos de predicciones INTRA e INTER con mayor eficiencia. En la Tabla # 2 se muestran algunas de las características de los perfiles utilizados en la codificación.

Tabla 2
 Características de los perfiles de codificación utilizados en FFMPEG

| Parámetros/Perfiles | <i>ultrafast</i> | <i>Medium</i> | <i>Placebo</i> |
|---------------------|------------------|---------------|----------------|
| CTU | 32 | 64 | 64 |
| <i>Min-cu-size</i> | 16 | 8 | 8 |
| <i>B-Frames</i> | 3 | 4 | 8 |
| <i>B-adapt</i> | 0 | 2 | 2 |
| <i>scenecut</i> | 0 | 40 | 40 |
| <i>Fast-intra</i> | Habilitado | Deshabilitado | Deshabilitado |
| Filtro SAO | Deshabilitado | Habilitado | Habilitado |
| <i>rdLevel</i> | 2 | 3 | 6 |

En la tabla # 3 es posible observar el *bitrate* obtenido en cada uno de los procesos de codificación por perfil de trabajo.

Tabla 3
Bitrate obtenido en cada muestra codificada en Kbps.

| Muestras/Perfiles | <i>ultrafast</i> | <i>Medium</i> | <i>Placebo</i> |
|---------------------|------------------|---------------|----------------|
| Resolución Estándar | 854,86 | 805,22 | 789,63 |
| Resolución FullHD | 2116,58 | 2095,93 | 1916,18 |
| Resolución 4K | 12615,08 | 11383,51 | 10361,37 |

Como se puede apreciar el comportamiento del *bitrate*, para cada muestra, es descendente con el cambio hacia un perfil de mayor complejidad de procesamiento. Las razones de este fenómeno radican en que el perfil *ultrafast* utiliza un tamaño de CTU de 32x32, con respecto a *medium* y *placebo* donde la CTU es de 64x64 y son capaces de aprovechar y comprimir, en mayor grado, las extensas zonas planas de las imágenes. Además de lo anterior expuesto, el perfil *ultrafast* no realiza un análisis exhaustivo para optimizar la razón de distorsión, ya que *rdlevel* tiene un valor de 2 y ello indica que solo realiza procesos de división y combinación de imágenes, omitiendo las predicciones basadas en el método SAD (*Sum of Absolute Differences*) que se utilizan a plenitud en el perfil *placebo*. Ello influye cuantitativamente en la eficiencia de la compresión y en el valor final del *bitrate* de la señal de video. Desde otro punto de vista analítico, es necesario hacer la observación del desempeño del códec h.265 en señales de video de altas resoluciones con respecto a las resoluciones menores. Bajo estas condiciones y considerando al perfil *ultrafast* como el tope teórico de máximo o mayor *bitrate* obtenido después de la compresión, la muestra 4K en el perfil *placebo* logra comprimir hasta un *bitrate* que representa el 82 % del obtenido en *ultrafast*. Para el caso de la muestra FullHD el perfil *placebo* logra reducir el *bitrate* hasta el 90 % del obtenido en el perfil *ultrafast*. En el caso de la muestra de resolución estándar la reducción lograda es del 92 %, realizando un análisis similar a los casos anteriores. Evidentemente, estas reducciones de *bitrate* demandan de un mayor tiempo de codificación. En la tabla # 4 se expresan los tiempos de codificación para cada una de las muestras en correspondencia con el perfil utilizado.

Tabla 4
Tiempos de codificación obtenidos en segundos.

| Muestras/Perfiles | <i>ultrafast</i> | <i>Medium</i> | <i>Placebo</i> |
|---------------------|------------------|---------------|----------------|
| Resolución Estándar | 3,62 | 15,13 | 550,57 |
| Resolución FullHD | 5,97 | 18,98 | 715,33 |
| Resolución 4K | 136,3 | 259,73 | 9150,97 |

En el caso de los tiempos de codificación se puede observar que los menores tiempos se obtienen para el perfil *ultrafast*. Primeramente, se debe aclarar que si el hecho de poseer CTU de 64x64 es una ventaja cuando se necesita un nivel de compresión alto, esto es una limitación a la hora de realizar el procesamiento paralelo de la secuencia de imágenes con el WPP. En el caso de *medium* y *placebo* utilizan la mitad de las filas de procesamiento paralelo que *ultrafast*, en todas las muestras (por ejemplo, en la muestra de resolución estándar *ultrafast* utiliza 68 filas mientras que *medium* y *placebo* 34). En *ultrafast* se encuentra habilitada la opción *fast-intra*, lo que implica que solo son chequeados entre 8 y 10 vectores en saltos de 5 o 6. La predicción intra-angular realizada se basa en la predicción de los vectores de mejor desempeño. En los perfiles *medium* y *placebo* se analizan los 33 vectores de predicción direccional. Además, en *ultrafast* no se realiza la colocación forzada de imágenes I (*scenecut=0*) por lo que el búffer de colocación anticipada (*Lookahead*) tiene un tamaño pequeño, incidiendo de forma determinante en la latencia del proceso de codificación. *Ultrafast* posee una estructura fija de GOP (*Group of Pictures*) con 3 imágenes tipo B en cada GOP. Ello implica que el nivel de esfuerzo del codificador para determinar la colocación de este tipo de imágenes es mínimo. En el caso de *medium* y *placebo* se utilizan algoritmos de Trellis para determinar la colocación de imágenes tipo B, lo que incide directamente en el monto total del proceso de codificación.

En el caso del PSNR los resultados obtenidos en el proceso de codificación son calculados por la herramienta FFMPEG y expresados explícitamente en cada una de las señales que componen el video digital. El valor tomado como resultado es el valor global del PSNR y se muestran en la Tabla # 5.

Tabla 5
Valores de PSNR por perfil de codificación

| Perfiles/Muestras | Definición estándar | FullHD | 4K |
|-------------------|---------------------|--------|----|
| | | | |

| | | | |
|-----------|-----------|-----------|-----------|
| ULTRAFAST | 34,992 dB | 36,177 dB | 38,989 dB |
| MEDIUM | 35,778 dB | 37,508 dB | 39,415 dB |
| PLACEBO | 35,886 dB | 38,177 dB | 39,581 dB |

Si se conoce que el valor de PSNR depende básicamente de la resolución de video y de la cantidad de bits por muestra, y que las tres muestras poseen aproximadamente la misma profundidad de bits, entonces las variaciones de los resultados obtenidos en esta métrica vienen dadas por las diferencias entre las resoluciones de video de las muestras codificadas, en gran medida. En el caso de la muestra SD con resolución estándar, los valores son los más cercanos al margen mínimo (30 dB). En las muestras de mayores resoluciones, los valores crecen hasta casi 40 dB, lo que pone de manifiesto sus potencialidades, teniendo en cuenta que se aprovecha la redundancia espacial y temporal durante la compresión.

Desde el punto de vista de los tres perfiles de desempeño seleccionados, se observa en la figura 5 que para el perfil *Ultrafast* se obtienen los menores valores de PSNR. Esto se debe a que este perfil limita la intensidad de la codificación para reducir la latencia durante el proceso, arriesgando en cierta medida la calidad de la compresión. En el perfil *Placebo* el valor es el mayor puesto que el codificador hace una compresión más profunda, es decir, utiliza la totalidad de las herramientas disponibles. Esto implica que el tiempo de codificación sea mayor. El perfil *Medium* realiza un compromiso entre intensidad de codificación y tiempo, para lograr valores aceptables de PSNR en un tiempo razonable.

Valores de PSNR obtenidos para cada muestra en los tres perfiles de desempeño de codificación H.265/HEVC

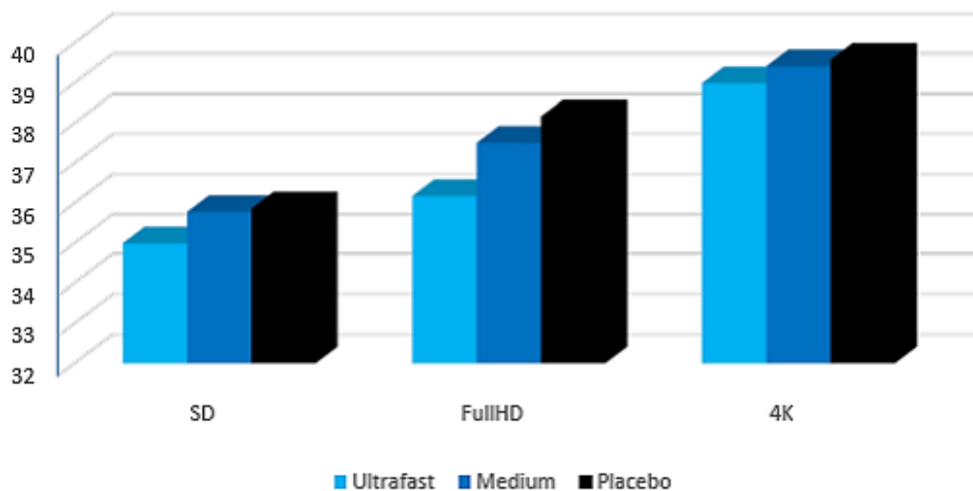


Figura 5

PSNR obtenido por resolución

En el caso del SSIM, al igual que el PSNR, FFMPEG brinda los valores de esta métrica. En la Tabla # 6 se aprecian los resultados obtenidos.

Tabla 6
 Valores de SSIM por perfil de codificación

| | | | |
|-------------------|---------------------|--------|----|
| Perfiles/Muestras | Definición estándar | FullHD | 4K |
|-------------------|---------------------|--------|----|

| | | | |
|-----------|--------|--------|--------|
| ULTRAFast | 0,8735 | 0,8846 | 0,9 |
| MEDIUM | 0,9171 | 0,9286 | 0,9408 |
| PLACEBO | 0,9726 | 0,9837 | 0,9907 |

Los valores de SSIM son menores que la unidad. Para el perfil *ultrafast* los valores son los menores de los obtenidos en los tres perfiles de desempeño de HEVC. Este resultado está justificado ya que el perfil *ultrafast* tiene deshabilitado el uso del filtro SAO y, en consecuencia, las distorsiones *banding* y *ringing* no son eliminadas. Los videos resultantes poseen distorsión con respecto a su muestra sin compresión correspondiente ya que el nivel de optimización de la razón de distorsión es el mínimo. Además, la predicción bidireccional utilizada por *ultrafast* emplea solamente tres imágenes de tipo B, ganando en velocidad de codificación, pero comprometiendo la semejanza estructural del video resultante. Otro factor que explica que los resultados de los valores de SSIM sean los menores para el perfil *ultrafast* es que este perfil emplea un tamaño de la CTU de 32x32 píxeles, o sea, el valor es la mitad del valor máximo establecido por HEVC, esto trae como consecuencia que, durante la codificación con este perfil, el codificador no aproveche al máximo la similitud que puede existir en las zonas planas de la imagen. En los perfiles *medium* y *placebo* los valores son mayores que 0.9100 y menores que 0.9910. Ambos perfiles utilizan tamaños de CTU de 64x64 píxeles, aprovechando al máximo la compresión en las zonas planas de las imágenes. Además, habilitan el uso del filtro SAO y consecuente a esto, se eliminan las distorsiones que se producen alrededor de los objetos durante la compresión. Las diferencias en los valores de SSIM resultante radican, en mayor medida, en la predicción bidireccional puesto que el perfil *medium* emplea cuatro imágenes de tipo B, y el perfil *placebo* emplea ocho, o sea, *placebo* está diseñado para explotar al máximo la predicción bidireccional, mientras que *medium* solo hace uso de esta predicción a un nivel intermedio. En la figura 6 se muestra este comportamiento.

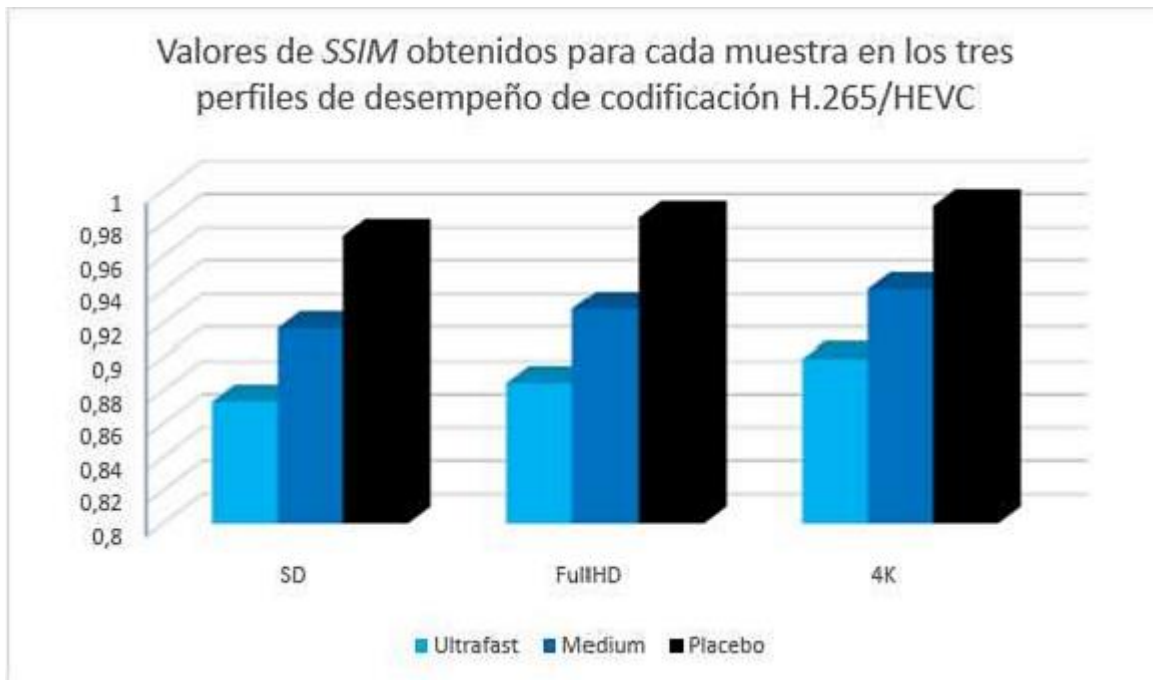


Figura 6

SSIM obtenido por resolución

4.- CONCLUSIONES

En este artículo se realizó la compresión eficiente de secuencias de video con diferentes resoluciones. Se utilizó un clúster de alto rendimiento y la herramienta de codificación FFMPEG, de software libre. A partir de los resultados obtenidos se llega a las siguientes conclusiones. Bajo estas condiciones de hardware, el perfil *placebo* ofrece los niveles más altos de

compresión al lograr los menores valores de *bitrate* en todas las muestras analizadas. Esto es posible sacrificando el tiempo de procesamiento a partir de la utilización intensiva de las técnicas y herramientas implementadas en el estándar HEVC/H.265. En adición, *placebo* posee los mayores valores de PSNR y SSIM, en todas las secuencias, destacando la muestra 4K con valores de 39,581 y 0,9907, respectivamente. Este perfil de trabajo pudiera ser el ideal para un proceso donde el video no requiera ser manejado en tiempo real y se pretenda maximizar la calidad de la imagen resultante, así como el ahorro de los recursos del canal de comunicaciones. El perfil *ultrafast* ofrece los menores tiempos de codificación a expensas de la calidad de la imagen y el *bitrate*, ya que aprovecha de manera eficiente el procesamiento paralelo y utiliza sólo herramientas indispensables para el proceso de codificación. El perfil *medium* ofrece un equilibrio entre tiempo de procesamiento y calidad del video codificado. Además, es preciso resaltar las potencialidades del códec H.265/HEVC para el trabajo con secuencias de video de altas resoluciones. La muestra 4K evidenció los mayores niveles de compresión de un perfil de trabajo a otro, comparado con las demás secuencias, al realizar una reducción del *bitrate* de un 18 % desde el perfil *ultrafast* hasta el *placebo*.

REFERENCIAS

1. ITU-T RH. 265 (04/13). Ser H Audiov Multimed Syst Infrastruct Audiov Serv Mov Video High Effic Video Coding [Internet]. Disponible en: <http://www.itu.int/rec/T-REC-H:265-201304>
2. Sullivan GJ, Ohm J, Han W-J, Wiegand T. Overview of the high efficiency video coding (HEVC) standard. IEEE Trans Circuits Syst Video Technol. 2012;22(12):1649–1668.
3. Sze V, Budagavi M, Sullivan GJ. High efficiency video coding (HEVC). Integr Circuit Syst Algorithms Archit Springer. 2014;39:40.
4. Molina PU. Técnicas de mejora de la eficiencia de codificación de vídeo [Tesis de Doctorado]. Universidad Politécnica de Valencia; 2016. Disponible en: <https://riunet.upv.es/handle/10251/59446>
5. Pourazad MT, Doutre C, Azimi M, Nasiopoulos P. HEVC: The new gold standard for video compression: How does HEVC compare with H. 264/AVC? IEEE Consum Electron Mag. 2012;1(3):36–46.
6. Grois D, Marpe D, Mulayoff A, Itzhaky B, Hadar O. Performance comparison of h. 265/mpeg-hevc, vp9, and h. 264/mpeg-avc encoders. En: Picture Coding Symposium (PCS), 2013. IEEE; 2013. p. 394–397.
7. Pastuszak G, Trochimiuk M. Algorithm and architecture design of the motion estimation for the H. 265/HEVC 4K-UHD encoder. J Real-Time Image Process. 2016;12(2):517–529.
8. Misra K, Segall A, Horowitz M, Xu S, Fuldseth A, Zhou M. An overview of tiles in HEVC. IEEE J Sel Top Signal Process. 2013;7(6):969–977.
9. Zhang S, Zhang X, Gao Z. Implementation and improvement of wavefront parallel processing for HEVC encoding on many-core platform. En: Multimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference on. IEEE; 2014. p. 1–6.
10. Hanhart P, Korshunov P, Ebrahimi T. Benchmarking of quality metrics on ultra-high definition video sequences. En: Digital Signal Processing (DSP), 2013 18th International Conference on. IEEE; 2013. p. 1–8.
11. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process. 2004;13(4):600–612.

AUTORES

Rafael Alejandro Olivera Solís. Ingeniero en Telecomunicaciones y Electrónica, Máster en Ciencias Telemáticas. Se desempeña como profesor asistente en el departamento de Electrónica y Telecomunicaciones de la Facultad de Ingeniería

Eléctrica de la Universidad Central “Marta Abreu” de las Villas. Es profesor principal de las asignaturas Radiopropagación y Líneas de Transmisión. Investiga en los temas de la compresión eficiente de señales de video con H.265/HEVC con aplicación en la TV digital terrestre, en la QoS de redes IP y redes WLAN. Se desempeña como vicedecano de Extensión, Comunicación e Informatización en dicha Facultad. rolivera@uclv.edu.cu

Yazdani López Pérez. Ingeniero en Telecomunicaciones y Electrónica. Especialista de la empresa Radiocuba en la Provincia de Cienfuegos. yazdani@cfg.radiocuba.cu



Los contenidos de la revista se distribuyen bajo una licencia Creative Commons Attribution-NonCommercial 3.0 Unported License