

Compresión eficiente de materiales audiovisuales utilizando H.265-HEVC

(Efficient video compression using H.265-HEVC)

Rafael Alejandro Olivera Solís^b, Yazdani López Pérez^a

^aUniversidad Central “Marta Abreu” de Las Villas

Resumen

HEVC/H.265 es el nuevo estándar de codificación de vídeo. 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 su calidad lo mayor posible, ha traído consigo la utilización de herramientas que permitan un proceso de codificación eficiente y eficaz. Este trabajo realiza un análisis del efecto que ejerce la variación de los parámetros de codificación sobre las métricas de compresión eficiente, utilizando herramientas de software libre y un clúster de alto rendimiento.

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

HEVC / H.265 is the new video encoding standard. Its main objective is to improve the levels of video compression. Video compression has played a key role in the development of multimedia applications. The need to reduce the bitrate, while maintaining its quality as much as possible, has brought with it the use of tools that allow an efficient and effective coding process. This work performs an analysis of the effect of the variation of the coding parameters on the efficient compression metrics, using free software tools and a high performance cluster.

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

Email addresses: ^brolivera@uclv.edu.cu (Rafael Alejandro Olivera Solís^b), ^aylperez@uclv.cu (Yazdani López Pérez^a).

1. Introducción

En la sociedad actual es innegable la importancia que tiene el vídeo 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) se ven afectadas por el limitado ancho de banda, así como por el reducido espectro de la señal. Por tanto, es necesario emplear técnicas de comprensión 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*), el formato de compresión de video de alta eficiencia. Este estándar fue desarrollado por ITU-T VCEG (*Video Coding Experts Group*) e ISO/IEC MPEG (*Moving Picture Experts Group*). Su objetivo principal fue mejorar la compresión de vídeo, en relación a los anteriores estándares (Sullivan et al., 2012), (ITU-T, 2013).

El nuevo estándar HEVC permite un importante avance en la compresión con respecto a sus predecesores (Sze et al., 2014). Las aplicaciones para HEVC no sólo cubrirán el espacio de los conocidos usos actuales y capacidades del video digital, sino que también incluirán el despliegue de nuevos servicios y la entrega de una calidad de video mejorada, como el despliegue de imágenes de Televisión de Ultra Alta Definición (UHDTV) y video con mayor rango dinámico, una gama más amplia de color representable y mayor precisión de representación de lo que se suele encontrar hoy en día Sze et al. (2014). 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.

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. 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*). En este artículo se analiza el efecto que ejerce la variación de los parámetros de codificación sobre las métricas de calidad de compresión eficiente. 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. Se describen los experimentos realizados y se muestran los resultados obtenidos.

2. Materiales y Métodos

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. Este se perfila como 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.

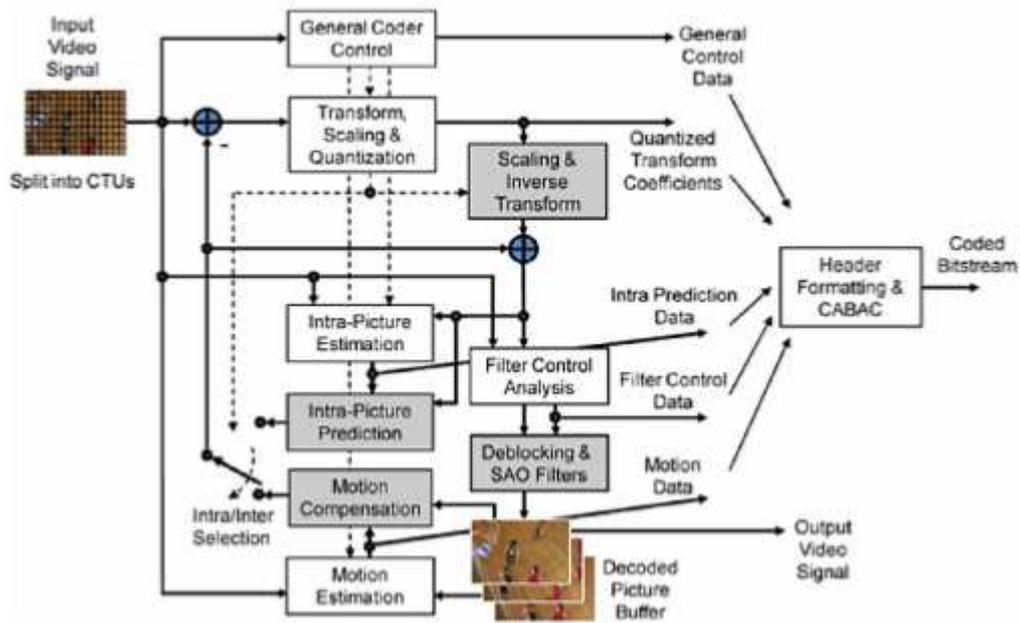


Figura 1. Diagrama de bloques del codificador de video HEVC (Molina, 2015).

HEVC/H. 265 mejora las herramientas existentes en los estándares previos, especialmente 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 UHDTV (*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 también 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.

1. 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 LxL 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 (Pourazad et al., 2012).

2. Unidad de Predicción: H.265 define una nueva unidad de predicción (PU) que pueden tener dimensiones desde 32x32 hasta 4x4, 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.
3. Predicción *Intra-Frame*: HEVC introduce un nuevo método de predicción *intra-frame*, el *intra-angular*, el cual 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 16x16, y 8 predictores para 8x8 y 4x4. 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 (Grois et al., 2013).
4. 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*, y 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 (Pastuszak and Trochimiuk, 2016).
5. Unidades de Transformación: Se define una nueva TU (*Transform Unit*). Se aplica de igual forma la DCT (*Discrete Cosine Transform*) con una ligera modificación con respecto a la utilizada en H.264, también permite bajo ciertas condiciones la utilización de la DST (*Discrete Sine Transform*). En el caso de H.265 se amplían los tamaños de transformación desde 4x4 a 32x32.
6. 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 (ITU-T, 2013).

Debido a la alta complejidad de los algoritmos introducidos por H.265, el estándar integra novedosas herramientas con el objetivo de frenar 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 vídeo que puede ser explotado en sistemas con arquitecturas de procesamiento paralelo siendo más flexibles que las *slices* de H.264 además de considerarse una tecnología menos compleja que FMO (*Flexible Macroblock Ordering*) (Caño Velasco, 2014).

WPP: Es una herramienta utilizada por HEVC en el procesamiento en paralelo. Cuando WPP 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.

• Métricas de Compresión Eficientes

Las primeras medidas objetivas de la calidad del video utilizadas son del tipo *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 la relación señal a ruido de pico (PSNR *Peak Signal to Noise Ratio*), definidas en las ecuaciones (1) y (2). Como se mencionó, estas métricas son del tipo FR (*Full Reference*), ya que requieren de la referencia completa para poder ser calculadas.

PSNR

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 vídeo, 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 (Martínez, 2016):

$$PSNR = 10 \log \left(\frac{MAX^2}{MSE} \right) = 20 \log \left(\frac{MAX}{\sqrt{MSE}} \right) \quad (1)$$

Donde MSE (*Mean Squared Error*, error cuadrático medio) 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.

SSIM

Esta medida de calidad visual objetiva, creada por Zhou Wang en 2004 (Martínez, 2016), nace con la intención de crear una medida objetiva más cercana al sistema visual humano que el PSNR.

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 se basó en 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 del SSIM de una imagen se realiza en ventanas de pequeño tamaño, normalmente de 8x8 píxeles, mediante la siguiente expresión:

$$SSIM(X, Y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (4)$$

$$c_1 = (0,01 \cdot MAX)^2 \quad (5)$$

$$c_2 = (0,03 \cdot MAX)^2, \quad (6)$$

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.

3. Resultados y discusión

La secuencia SD consta de 600 fotogramas, con una resolución de 704x576 píxeles y submuestreo en los planos de diferencia de color de 4:2:0.



Figura 2. Secuencia SD.

La secuencia *FullHD* consta de 217 fotogramas, con una resolución de 1920x1080 píxeles y sub-muestreo en los planos de diferencia de color 4:2:0.



Figura 3. Secuencia *FullHD*.

La secuencia 4K está formada por 600 fotogramas, con resolución de 4096x2160 píxeles y sub-muestreo en planos de diferencia de color de 4:2:0.



Figura 4. Secuencia 4K.

Las tres muestras poseen resoluciones variables: *Standard Definition*, *Full HD* y *4KHD*. En este sentido, el códec H.265 analiza las posibilidades de que existan grandes zonas en la imagen con características similares (Cuadro 1), en mayor medida para las resoluciones 1080p y 2160p, y explota los tipos de predicciones INTRA e INTER. En los tres casos, los dos modos de predicciones mencionados se tienen en cuenta para las codificaciones, puesto que las características visuales de las tres muestras evidencian que se puede lograr un aprovechamiento de la redundancia tanto espacial como temporal, existentes en las secuencias.

Se observa en el Cuadro 2 que, para las tres muestras de videos, los valores de PSNR se encuentran dentro del rango típico de 30 dB a 50 dB para profundidades de 8 y 10 bits, en una compresión con pérdidas.

Cuadro 1

Características de las muestras de video sin compresión

Parámetros/Muestras	<i>SD</i>	<i>FullHD</i>	<i>4K</i>
Profundidad de bits	8 bits	8 bits	10 bits
Bitrate	291 964 kbps	622 080 kbps	4 150 263 kbps
Tamaño del archivo	348 MB	643 MB	14.8 GB
Formato	YUV4MPEG2	YUV4MPEG2	YUV4MPEG2
Relación de Aspecto	4:3	16:9	1.896:1
Frame rate (fps)	60.000	25.000	60.000
Número Total de Fotogramas	600	217	600
Duración	10 segundos	9 segundos	10 segundos
Submuestreo de Color	4:2:0 , S=1.5	4:2:0 , S=1.5	4:2:0 , S=1.5
INTRA	sí	sí	sí
INTER	sí	sí	sí

Cuadro 2

Valores de PSNR por perfil de codificación.

Perfiles/Muestras	<i>SD</i>	<i>FullHD</i>	<i>4K</i>
Ultrafast	34.992 dB	36.177 dB	38.989 dB
Medium	35.778 dB	37.508 dB	39.415 dB
Placebo	35.866 dB	38.166 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. Para este caso la compresión fue la menos eficiente debido a que el video tiene la menor de las resoluciones, y el estándar HEVC es más eficiente en el procesamiento de videos de altas resoluciones, dígame 1080p, 2160p, etc. 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 debido 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 *medium*, el valor de la métrica es mayor puesto que el codificador hace una compresión más profunda, por tanto, existe una relación de proporcionalidad directa entre la intensidad de codificación en los perfiles de desempeño (más a la derecha, mayor intensidad) con los valores de PSNR obtenidos.

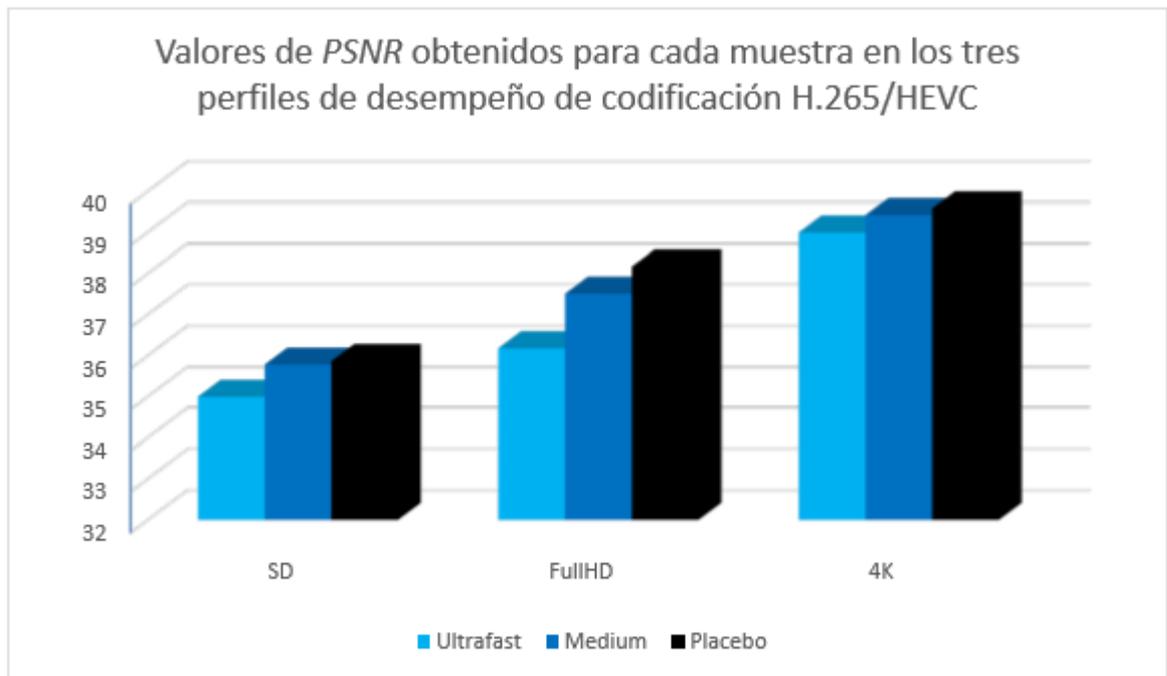


Figura 5. Resultados de PSNR obtenidos.

Cuadro 3

Valores de SSIM por perfil de codificación.

Perfiles/Muestras	<i>SD</i>	<i>FullHD</i>	<i>4K</i>
Ultrafast	0.8735 dB	0.8847 dB	0.8991 dB
Medium	0.9171 dB	0.9286 dB	0.9408 dB
Placebo	0.9726 dB	0.9837 dB	0.9907 dB

Tal y como se observa en el Cuadro 3, los valores de SSIM son menores que la unidad, pero relativamente aproximados y superiores al valor 0.8700, lo cual representa una compresión altamente eficiente, donde los videos resultantes poseen baja distorsión estructural en comparación con las muestras iniciales sin compresión y, por tanto, mayor similitud. Para el perfil *ultrafast* se observa en la Figura 6 que los valores son los menores de los obtenidos en los tres perfiles de desempeño de HEVC. Este resultado está justificado porque el perfil *ultrafast* tiene deshabilitado el uso del filtro SAO y, en consecuencia, las distorsiones banding y ringing no son eliminadas, por tanto, la codificación no es la más eficiente, y los videos resultantes poseen cierta distorsión estructural con respecto a su muestra sin compresión correspondiente. 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, esto demuestra que el proceso de codificación se realizó con mayor intensidad. Ambos perfiles utilizan

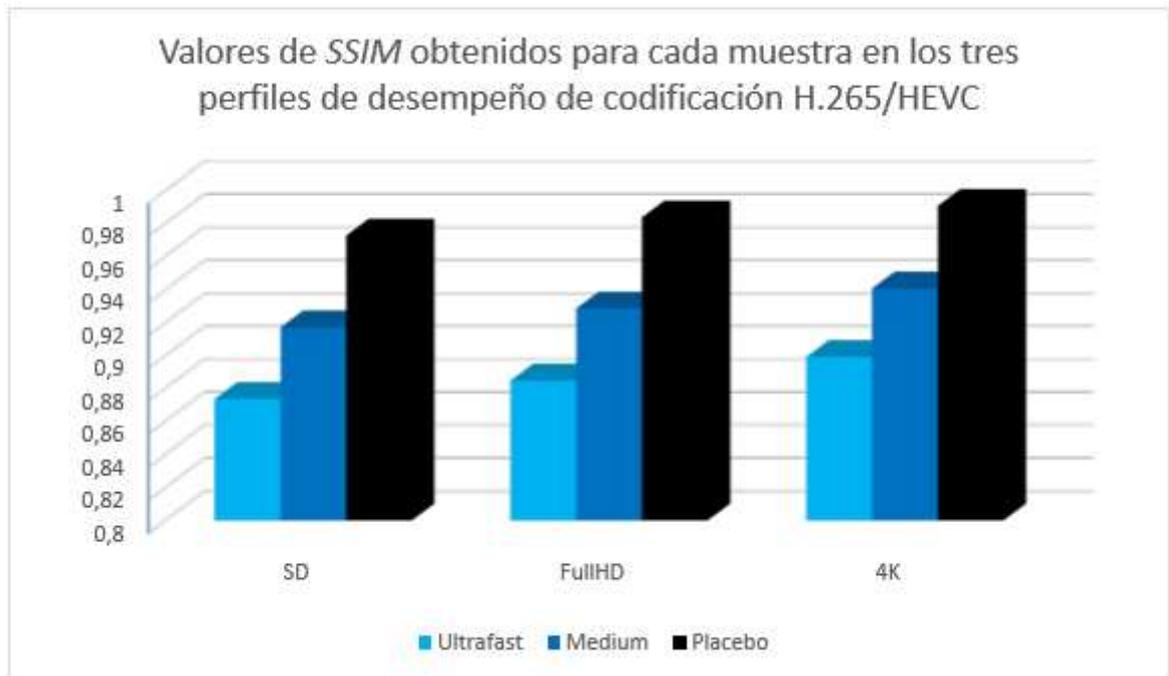


Figura 6. Resultados de SSIM obtenidos.

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 equilibrado. Es meritorio aclarar que todos los valores obtenidos de PSNR y SSIM son valores promedios y absolutos respectivamente ya que cada componente de la señal de video (Y,U,V) aporta su valor independiente de cada una de las métricas.

4. Conclusiones

La variación de los parámetros de cada uno de los perfiles de desempeño de H.265/HEVC conllevan a una alteración en la eficiencia de codificación y en los valores de las métricas objetivas de calidad visual. Al disminuir la profundidad e intensidad de codificación, los valores de las métricas disminuyen, y por ende, disminuye la eficiencia de codificación. Si por el contrario se intensifica la codificación, los valores de las métricas de calidad visual son mayores y resulta una mayor eficiencia durante el proceso. El estándar HEVC proporciona una reducción elevada del bitrate del video a transmitir por el canal de TV, posibilitando un mayor aprovechamiento de dicho canal al poder multiplexar en un mismo canal de transmisión mayor número de programas con la mayor calidad posible. Las métricas de calidad visual constituyen las herramientas más confiables para evaluar la eficiencia de la codificación. Si se desea mejor correlación con el sistema visual humano, la mayor eficiencia se obtiene en los casos donde el valor de

SSIM se aproxima más a uno. Por otra parte, si se desea determinar la mejor compresión, se selecciona la codificación cuyo valor de PSNR es mayor. El perfil *ultrafast* garantiza una codificación en tiempo real pero devuelve valores de las métricas de calidad visual menores, por tanto, una menor eficiencia. Sin embargo, el perfil *placebo* introduce cierta latencia en el proceso de codificación, pero devuelve mayores valores de las métricas objetivas de calidad visual, resultando una mayor eficiencia en el proceso. El perfil *medium* garantiza un equilibrio entre los valores de las métricas de calidad visual, buena eficiencia de codificación y una latencia aceptable en el proceso.

5. Referencias bibliográficas

Referencias

- Caño Velasco, J. P., 2014. Caracterización de un decodificador hevc ejecutándose en un dsp.
- Grois, D., Marpe, D., Mulyoff, A., Itzhaky, B., Hadar, O., 2013. Performance comparison of h. 265/mpeg-hevc, vp9, and h. 264/mpeg-avc encoders. In: Picture Coding Symposium (PCS), 2013. IEEE, pp. 394–397.
- ITU-T, 2013. 265 (04/13). Series H: Audiovisual and Multimedia Systems, Infrastructure of audiovisual services–Coding of Moving Video, High Efficiency Video Coding, Online: <http://www.itu.int/rec/T-REC-H,265-201304>.
- Martínez, P., 2016. Compresión eficiente de materiales audiovisuales utilizando h.265-hevc. Trabajo de diploma, Universidad Central "Marta Abreu" de Las Villas.
- Molina, P. U., 2015. Técnicas de mejora de la eficiencia de codificación de video. Master's thesis, Universidad Politécnica de Valencia.
- Pastuszak, G., Trochimiuk, M., 2016. Algorithm and architecture design of the motion estimation for the h. 265/hevc 4k-uhd encoder. *Journal of Real-Time Image Processing* 12 (2), 517–529.
- Pourazad, M. T., Doutre, C., Azimi, M., Nasiopoulos, P., 2012. Hevc: The new gold standard for video compression: How does hevc compare with h. 264/avc? *IEEE consumer electronics magazine* 1 (3), 36–46.
- Sullivan, G. J., Ohm, J., Han, W.-J., Wiegand, T., 2012. Overview of the high efficiency video coding (hevc) standard. *IEEE Transactions on circuits and systems for video technology* 22 (12), 1649–1668.
- Sze, V., Budagavi, M., Sullivan, G. J., 2014. High efficiency video coding (hevc). *Integrated Circuit and Systems, Algorithms and Architectures*. Springer, 1–375.