

Análisis perceptual de las herramientas de codificación en el estándar HEVC

Javier Ruiz Atencia¹, Otoniel López Granado¹, Manuel Pérez Malumbres¹, Miguel Onofre Martínez-Rach¹ y Héctor Migallón¹

Resumen— Con la llegada de cada nuevo estándar de vídeo se añaden nuevas herramientas de codificación, o bien se evolucionan las ya existentes, con el objetivo de mejorar el rendimiento y la calidad con respecto a sus predecesores. La gran mayoría de estas herramientas de codificación utilizan la métrica PSNR como medida de distorsión. Esta métrica, basada en el error cuadrático medio, no tiene en cuenta el sistema visual humano, por lo que no es posible determinar si las herramientas obtienen resultados perceptualmente óptimos. Con objeto de profundizar en este punto, hemos desarrollado un estudio detallado sobre la respuesta perceptual de las diferentes herramientas de codificación del estándar HEVC. Para medir esta respuesta, hemos utilizado algunas de las métricas perceptuales más populares en la actualidad. Las conclusiones de este estudio ayudarán a determinar el conjunto de herramientas de codificación que proporcionen, en general, la mejor respuesta perceptual.

Palabras clave— HEVC, codificación perceptual, filtro deblocking, filtro SAO, transform skip, RDOQ, CSF, métricas perceptuales.

I. INTRODUCCIÓN

HIGH Efficiency Video Coding (HEVC) es uno de los últimos estándares de codificación de vídeo desarrollado por las organizaciones de estandarización Joint Collaborative Team on Video Coding (JCT-VC) del ITU-T Video Coding Experts Group (VCEG) junto con el ISO/IEC Moving Picture Experts Group (MPEG). Su primera versión se hizo pública en 2013 y, hoy en día, sigue en continuo desarrollo.

El objetivo principal del estándar HEVC era el de reducir en hasta un 50 % el bit rate manteniendo la misma calidad subjetiva con respecto a su predecesor, el estándar H.264/AVC, sin incrementar la complejidad del codificador. Pa-

ra conseguir este objetivo, el estándar HEVC incorpora numerosas técnicas de codificación que intentan reducir el bit rate sin incrementar la distorsión. Muchas de estas técnicas son heredadas del estándar H.264/AVC, mientras otras son de nueva creación, como por ejemplo el particionado Quad-tree o el filtro SAO (Sample Adaptive Offset).

Algunas herramientas de codificación del estándar HEVC incluyen aproximaciones al comportamiento del Sistema Visual Humano (HVS), con el fin de tener en cuenta la calidad subjetiva percibida por el ojo humano. En particular, HEVC proporciona la herramienta de codificación ScalingList (SCL), que aplica una cuantización no uniforme a los coeficientes transformados, dependiendo de la sensibilidad al contraste asociada a sus frecuencias. La idea principal es aumentar el valor de la cuantización en las regiones donde el HVS es menos sensible [1]. Otros estudios incluyen técnicas de enmascaramiento por luminancia en secuencias de vídeo de alto rango dinámico (HDR) [2] [3], logrando una importante reducción del bit rate sin perjudicar la calidad perceptual.

Aunque las técnicas empleadas durante el proceso de codificación en el estándar HEVC han demostrado ser capaces de reducir el bit rate, no se puede garantizar que sean óptimas desde el punto de vista perceptual. Durante el desarrollo de estas técnicas, la métrica empleada para medir la distorsión ha sido la PSNR (Peak Signal-to-Noise Ratio). Esta métrica proporciona un valor de calidad basándose en las diferencias entre los píxeles de las imágenes originales y reconstruidas. Debido a esto, es sabido que este tipo de métricas no reflejan con exactitud la evaluación perceptual de calidad [4][5][6][7][8].

¹Dpto. de Ingeniería de Computadores, Universidad Miguel Hernández de Elche, e-mail: {javier.ruiza,otoniell,mels,hmigallon,mmrach}@umh.es.

Por tanto, para evaluar adecuadamente el rendimiento perceptual de las herramientas de codificación del estándar HEVC, es necesario emplear otras métricas que proporcionen valores de calidad con un alto grado de correlación con la calidad percibida por los humanos. De este modo, nos aseguraremos de que las diferentes herramientas de codificación siempre se evalúen para maximizar la respuesta perceptual del codificador, evitando, en la medida de lo posible, el uso de pruebas subjetivas (como por ejemplo estudios DMOS), más costosas de realizar.

En este estudio, hemos analizado el rendimiento perceptual de las principales herramientas de codificación que distorsionan la reconstrucción de las imágenes y que se encuentran incluidas en el software de referencia HEVC Test Model (HM). Para ello, hemos analizado la respuesta perceptual de un conjunto variado de secuencias de vídeo codificadas con diferentes parámetros de configuración, intentando determinar el conjunto de herramientas que maximizan la calidad perceptual. Cada configuración determina qué herramientas de codificación se utilizan, por lo que no solo hemos analizado su contribución a la respuesta perceptual de forma individual sino también al combinarlas con otras herramientas de codificación.

Para medir la calidad de las diferentes codificaciones hemos empleado un conjunto de métricas de calidad objetivas utilizadas ampliamente por la comunidad científica, desarrolladas para tener en cuenta el sistema visual humano, así como la recientemente popular métrica VMAF (Video Multimethod Assessment Fusion), desarrollada por Netflix [9].

El resto del artículo se estructura de la siguiente manera. En la Sección II se presentan una visión general de las herramientas de codificación del estándar HEVC analizadas. En la Sección III explicamos la metodología utilizada en este estudio. La Sección IV muestra los resultados experimentales, mientras que la Sección V proporcionamos un breve análisis de los resultados obtenidos. Por último, en la Sección VI se resumen las conclusiones de este estudio y se indican algunas líneas de investiga-

ción futuras.

II. HERRAMIENTAS DE CODIFICACIÓN HEVC

El estándar HEVC incluye numerosos parámetros de configuración que permiten configurar, habilitar o deshabilitar determinadas herramientas de codificación con tal de mejorar la calidad de las reconstrucciones, reducir el peso del bit stream o simplificar la complejidad del codificador. Sin embargo, la modificación de estos parámetros puede afectar a la calidad perceptual de las imágenes.

En este trabajo, nos hemos centrado en evaluar el impacto en la calidad perceptual del codificador HEVC al habilitar o deshabilitar las siguientes herramientas de codificación: *ScalingList*, *filtro de Deblocking*, *filtro SAO*, *Rate-Distortion Optimized Quantization* y *Transform Skip*.

A. *ScalingList*

El sistema visual humano (HVS) no es capaz de detectar todas las frecuencias espaciales con la misma precisión [10]. Numerosos estudios se han encargado de caracterizar lo que se conoce como la función de sensibilidad al contraste (Contrast Sensitivity Function, CSF) a partir de mediciones del umbral de contraste en humanos para diferentes frecuencias espaciales [11][12][13], llegando a la conclusión de que el ojo humano tiene menor sensibilidad en las altas y bajas frecuencias que en las frecuencias intermedias.

Una versión de esta CSF se encuentra implementada en la etapa de cuantización del estándar HEVC, y puede ser habilitada mediante el parámetro *ScalingList*. Por defecto, el codificador aplica una cuantización uniforme para todos los coeficientes transformados (*ScalingList* = 0), es decir, no tiene en cuenta la función de sensibilidad al contraste. Al habilitar este parámetro (*ScalingList* = 1), se utiliza una matriz de cuantización no uniformes predefinidas en el estándar. Esta matriz, como la mostrada en la Figura 1-b, define un peso adicional a cada coeficiente que varía en función de su posición [14].

El resultado del estudio llevado a cabo por [15] demuestra que, en promedio, el uso de matrices de cuantización no uniformes proporcionan mejores resultados en cuanto a calidad subjetiva.

$\begin{bmatrix} 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \end{bmatrix}$	$\begin{bmatrix} 16 & 16 & 16 & 16 & 17 & 18 & 21 & 24 \\ 16 & 16 & 16 & 16 & 17 & 19 & 22 & 25 \\ 16 & 16 & 17 & 18 & 20 & 22 & 25 & 29 \\ 16 & 16 & 18 & 21 & 24 & 27 & 31 & 36 \\ 17 & 17 & 20 & 24 & 30 & 35 & 41 & 47 \\ 18 & 19 & 22 & 27 & 35 & 44 & 54 & 65 \\ 21 & 22 & 25 & 31 & 41 & 54 & 70 & 88 \\ 24 & 25 & 29 & 36 & 47 & 65 & 88 & 115 \end{bmatrix}$
(a)	(b)

Fig. 1: Matrices de cuantización 8×8 por defecto para (a) ScalingList = 0 y (b) ScalingList = 1.

B. Filtro de Deblocking

El filtro de Deblocking suaviza la transición entre los bloques en los que se divide cada fotograma durante el proceso de codificación. Se utiliza después de la reconstrucción del bloque, pero su implementación se realiza dentro del bucle de codificación (filtro In-loop), de esta forma los bloques reconstruidos y filtrados son tomados como referencia para la predicción de bloques adyacentes.

En HEVC, el codificador puede elegir adaptativamente entre dos niveles de filtrado (normal o strong), o no aplicar filtro en función de los bloques adyacentes y un determinado umbral [16]. Como ejemplo, se muestra la Figura 2, donde se compara el uso del filtro de deblocking en un fragmento ampliado de la secuencia BlowingBubbles. Como se puede ver, los artefactos de bloques se minimizan cuando el filtro se encuentra habilitado, sin embargo, algunas regiones que no deberían ser filtradas son suavizadas por el filtro.

En [16], los autores sostienen que el filtro de deblocking incrementa tanto la calidad objetiva como la subjetiva de las secuencias de vídeo codificadas.

C. Filtro SAO

El Sample Adaptive Offset (SAO) es un nuevo filtro incorporado al estándar HEVC. Se utiliza después del filtro de deblocking, junto con el cual forman la etapa de filtrado In-loop. El principal objetivo del filtro consiste en reducir la distorsión en el interior de los bloques



Fig. 2: Ejemplo de uso del filtro de deblocking en un frame de BlowingBubbles codificado con QP=37: (a) DB deshabilitado, (b) DB habilitado.

producida por el uso de transformadas de gran tamaño (16×16 y 32×32). Para ello, se categorizan los bloques en dos tipos, Edge Offset (EO) y Band Offset (BO), y se obtiene un valor de offset a aplicar a los píxeles de cada categoría. El índice de las distintas categorías y su offset son transmitidos a través del bit stream, mientras que la clasificación se realiza tanto en el codificador como en el decodificador, con el objetivo de reducir la información a transmitir [17].

En [17], los autores argumentan que el uso del filtro SAO puede suponer una ganancia promedio de 3,5% en BD-Rate [18] para la métrica objetiva PSNR. En cuanto a la calidad subjetiva, los autores determinan que, basándose en experimentos subjetivos llevados a cabo por ellos mismos, se percibe en general una mejora de calidad perceptual, siendo esta mejora mayor en secuencias de vídeo sintéticas, como puede verse en la Figura 3, donde el filtro SAO incrementa significativamente la calidad visual al reducir los artefactos de anillo próximos a los bordes (ringing artifacts).

D. Rate-distortion optimized quantization (RDOQ)

El estándar HEVC incorpora como método de cuantización por defecto el algoritmo RDOQ (Rate Distortion Optimized Quantization). Este método cuantiza los coeficientes de forma individual tratando de minimizar la función de coste R/D. El objetivo del RDOQ es encontrar un conjunto de coeficientes cuantizados óptimo para obtener un compromiso ade-

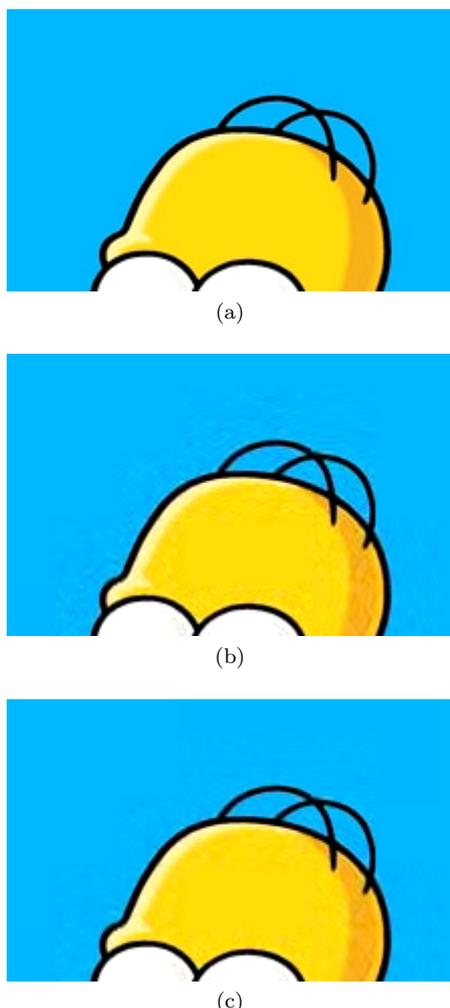


Fig. 3: Ejemplo de uso del filtro SAO en un frame sintético codificado con QP=32: (a) Original sin comprimir, (b) SAO deshabilitado, (c) SAO habilitado.

cuado entre la tasa de bits y la distorsión inducida por la cuantización. En esencia, el algoritmo modifica los valores predeterminados de las tablas de cuantización en función del rendimiento R/D [19].

El software de referencia HM incorpora también el parámetro RDOQTS, el cual habilita el algoritmo RDOQ cuando los bloques se marcan como Transform Skip, es decir, cuando se salta la etapa de transformación (ver Subsección II-E).

Hay que señalar que, siendo RDOQ un método eficaz para aumentar el rendimiento R/D,

en [20] los autores afirman que la optimización de la calidad basada en la métrica PSNR es perceptualmente insignificante desde el punto de vista en que un observador humano interpreta la calidad de un vídeo comprimido.

E. Transform Skip

El parámetro Transform Skip del estándar HEVC permite al codificador la opción de saltarse la etapa de transformación en bloques de tamaño 4×4 . De este modo, el error de predicción es directamente codificado en el dominio del píxel. Este algoritmo tiene una baja influencia en cuanto a rendimiento de codificación para las secuencias de clase A a E, sin embargo, para la clase F, se tiene de media una ganancia del 7,3% de términos de rendimiento de codificación (PSRN) [21].

A su vez, el software de referencia HM incluye un parámetro adicional, llamado Transform Skip Fast, el cual reduce la complejidad del algoritmo y por tanto acelera este proceso, a cambio de reducir su rendimiento en comparación con el algoritmo completo.

III. MÉTODOS Y PROCEDIMIENTOS

Para la elaboración de este estudio hemos seguido las directrices establecidas en el documento Common Test Condition [22]. Este documento define un marco normativo que establece un conjunto de secuencias y una configuración base para el software de referencia HM, cuya versión utilizada en este estudio ha sido la 16.20 [23]. En la Tabla I se muestra los valores por defecto de los parámetros a analizar. Las secuencias de vídeo se clasifican en seis categorías o clases (A-F). Cada clase determina una resolución o contenido determinado, por ejemplo, las clases A y B tienen una resolución elevada, la clase D tienen una resolución muy baja, mientras que la clase F tiene secuencias de vídeo sintéticas o generadas por ordenador. En la Tabla II se muestra el listado de secuencias y sus especificaciones más relevantes.

En este estudio nos hemos centrado exclusivamente en el modo de codificación All Intra Main (AI Main), es decir, cada frame se codifica con la información propia del frame (codificación espacial), sin tener en cuenta frames an-

Tabla I: Valor por defecto de los parámetros analizados

Parámetro	Valor
QP	22, 27, 32, 37, 42
ScalingList	0
LoopFilterDisable	0
SAO	1
RDOQ	1
RDOQTS	1
TransformSkip	1
TransformSkipFast	1

Tabla II: Especificaciones de las secuencias de test

Clase	Secuencia	Resolución	Frames	Tasa de frames	Prof. de Bit
A	Traffic	2560x1600	150	30	8
	PeopleOnStreet		150	30	8
	Nebuta		300	60	10
	SteamLocomotive		300	60	10
B	Kimono	1920x1080	240	24	8
	ParkScene		240	24	8
	Cactus		500	50	8
	BQTerrace		600	60	8
	BasketballDrive		500	50	8
C	RaceHorses	832x480	300	30	8
	BQMall		600	60	8
	PartyScene		500	50	8
	BasketballDrill		500	50	8
D	RaceHorses	416x240	300	30	8
	BQSquare		600	60	8
	BlowingBubbles		500	50	8
	BasketballPass		500	50	8
E	FourPeople	1280x720	600	60	8
	Johnny		600	60	8
	KristenAndSara		600	60	8
F	BasketballDrillText	832x480	500	50	8
	ChinaSpeed	1024x768	500	30	8
	SlideEditing	1280x720	300	30	8
	SlideShow	1280x720	500	20	8

teriores o posteriores (codificación temporal). La elección única de este modo viene dada por la limitación de la mayoría de las métricas perceptuales de no tener en cuenta la codificación temporal, por lo que no captan los artefactos producidos por este tipo de codificaciones.

Para medir el rendimiento rate-distortion (R/D) de las diferentes configuraciones del codificador, hemos utilizado el método Bjøntegard Delta Rate (BD-Rate) [18] en las distintas métricas perceptuales. Los valores del parámetro de cuantización (QP) seleccionados han sido 22, 27, 32, 37 y 42. Este último valor no aparece en el documento normativo del HEVC [22], sin embargo, se ha decidido añadirlo para incrementar el rango dinámico de las curvas R/D, y así aumentar la precisión

de los cálculos BD-Rate.

La evaluación de la influencia de los parámetros descritos en la Sección II se ha llevado a cabo codificando el conjunto de secuencias, habilitando y deshabilitando cada uno de los parámetros, obteniendo así un total de 32 configuraciones distintas. Debido al elevado número de codificaciones a realizar, se ha descartado el uso de métricas subjetivas como los test DMOS. En su lugar, hemos utilizado las siguientes métricas objetivas: SSIM, MS-SSIM, VIF y PSNR-HVS-M y VMAF.

Las métricas SSIM (Structural Similarity) [7] y MS-SSIM (Multi-Scale SSIM) [6] se basan en la hipótesis de que el sistema visual humano está altamente adaptado a extraer información estructural de las imágenes. Ambas métricas tienen en consideración la luminancia, el contraste y la información estructural de las escenas, mientras que la MS-SSIM también tiene en cuenta la escala.

La métrica VIF (Visual Information Fidelity) [24] se basa en el modelo llamado Natural Scene Statistics (NSS), que sostiene que las imágenes de alta calidad tienen propiedades estadísticas medibles. Además, cuenta también con un modelo de degradación de imagen y componentes del HVS para obtener la puntuación de calidad.

La métrica PSNR-HVS-M [25] está basada en la PSNR, pero incorpora la función de sensibilidad al contraste (CSF) y el efecto del enmascaramiento por contraste.

La métrica perceptual más reciente utilizada en este estudio es la VMAF, desarrollada por Netflix [9]. A diferencia del resto, la VMAF utiliza técnicas de aprendizaje automático para estimar la puntuación que se obtendría mediante la realización de pruebas subjetivas. Para ello, esta métrica ha sido entrenada con inputs de pruebas subjetivas DMOS, así como tres métricas objetivas: VIF, DLM (Detail Loss Measure) [26] y TI (Temporal Perceptual Information) [27]. Según diversos estudios, la métrica VMAF está fuertemente correlacionada con los valores perceptuales obtenidos con estudios MOS y DMOS [28][29][30].

Como se comentó anteriormente, dado que ninguna de las métricas perceptuales excepto

la VMAF incorpora medidas de distorsión temporal, para una justa comparación de resultados entre las diferentes métricas se ha decidido eliminar el factor temporal de la VMAF. Para ello, se han obtenido las puntuaciones para cada frame de las secuencias de vídeo de forma individual, tal y como se hace para el resto de métricas.

Por último, hay que mencionar que se ha llevado también a cabo un análisis de tiempos para cada uno de los parámetros de configuración, con el objetivo de determinar su complejidad media. Teniendo en cuenta tanto los resultados de las métricas perceptuales como la complejidad, podemos proponer la configuración de codificación adecuada que proporcione buenos resultados perceptuales con una complejidad media balanceada.

IV. RESULTADOS EXPERIMENTALES

En esta sección mostraremos los valores BD-Rate, promediados por clases, obtenidos tras codificar las 32 diferentes configuraciones del codificador descritas en la sección anterior. Estos resultados se muestran en las Tablas III-A a III-F, donde para cada tabla se muestra el rendimiento R/D de las diferentes configuraciones medido para cada métrica perceptual.

La primera columna de las tablas identifica cada una de las configuraciones de codificación. Las siguientes cinco columnas muestran el estado (habilitado o deshabilitado) de cada parámetro de configuración: SCL corresponde al parámetro *ScalingList*; SAO corresponde al filtro SAO; DB corresponde al filtro de deblocking; RDOQ incluye ambos parámetros RDOQ y RDOQTS; y TrSk corresponde a los parámetros *TransformSkip* y *TransformSkipFast*. Las puntuaciones de calidad obtenidas se expresan en porcentaje de reducción o incremento de BD-Rate.

Para facilitar la lectura de las tablas, hemos destacado en verde los valores negativos, que corresponden con ganancias en BD-Rate con respecto a la configuración por defecto (fila #1 de cada tabla). Las tablas individuales para cada secuencia de vídeo se pueden consultar a través de la web del grupo de investigación GATCOM [31].

En las siguientes subsecciones, describiremos los resultados obtenidos mostrando el comportamiento perceptual de cada herramienta de codificación en estudio, mientras que en la Sección V, se ofrecerá un análisis más generalista de estos resultados.

A. *ScalingList*

Por defecto, este parámetro se encuentra deshabilitado, por lo que hemos analizado la influencia perceptual al habilitarlo. Los resultados muestran que habilitar el parámetro SCL es perceptualmente beneficioso en prácticamente todas las configuraciones analizadas. Esto se puede ver fácilmente comparando las filas #1 a #16 (SCL deshabilitado) frente a las filas #17 a #32 (SCL habilitado), donde en este segundo grupo el valor de BD-Rate es generalmente menor que en el primer grupo para todas las métricas utilizadas. Si comparamos el impacto de esta herramienta con respecto a la configuración por defecto (fila #1), al habilitarla las métricas reportan ganancias reales (valores BD-Rate negativos) en cada una de las clases de secuencias de vídeo.

Con respecto a la complejidad de esta herramienta de codificación, al habilitarla se tiene entre un 3,47% y un 8,44% de incremento de tiempos de codificación, en función de la QP utilizada (ver Tabla IV).

B. *Filtro de Deblocking*

El filtro de deblocking minimiza el efecto de bloques producido por el modo de particionar las imágenes durante el proceso de codificación. Cuando se deshabilita el filtro, este efecto se vuelve mucho más visible, sobre todo con valores de QP elevados.

Como puede verse en las tablas III-A a III-F, en la mayoría de casos las métricas perceptuales dan mayor rendimiento perceptual a las configuraciones con el filtro de deblocking habilitado que con respecto a las mismas configuraciones con el filtro deshabilitado. Aunque este comportamiento varía en función de la clase de secuencia, se puede observar como las métricas SSIM, MS-SSIM y PSNR-HVS-M tienen de media pérdidas de BD-Rate de importantes (1,2% hasta un 2,5%). Sin embargo, las

Tabla III-A: Rendimiento promedio de codificación [% BD-Rate] para la Clase A

#	SCL	SAO	DB	RDOQ	TrSk	SSIM	MS-SSIM	VMAF	VIF	PSNR HVS-M
1	0	1	1	1	1	0	0	0	0	0
2	0	1	1	1	0	-0,03	-0,03	0,06	-0,04	-0,04
3	0	1	1	0	1	5,6	2,29	11,93	3,84	5,38
4	0	1	1	0	0	5,54	2,24	11,95	3,8	5,34
5	0	1	0	1	1	0,26	1,03	1,32	-0,27	1,51
6	0	1	0	1	0	0,24	1	1,38	-0,31	1,46
7	0	1	0	0	1	5,91	3,46	13,48	3,61	7,05
8	0	1	0	0	0	5,86	3,42	13,51	3,57	7
9	0	0	1	1	1	0,18	0,21	-2,08	0,03	0,26
10	0	0	1	1	0	0,15	0,19	-2,03	-0,01	0,21
11	0	0	1	0	1	5,78	2,47	10,02	3,82	5,61
12	0	0	1	0	0	5,73	2,43	10,03	3,8	5,56
13	0	0	0	1	1	0,38	1,56	-3,32	-0,47	3,05
14	0	0	0	1	0	0,35	1,53	-3,27	-0,51	2,99
15	0	0	0	0	1	6,01	4,08	8,68	3,28	8,65
16	0	0	0	0	0	5,95	4,03	8,69	3,24	8,59
17	1	1	1	1	1	-0,38	-0,2	-0,18	-0,34	-0,63
18	1	1	1	1	0	-0,44	-0,24	-0,13	-0,38	-0,68
19	1	1	1	0	1	4,28	1,41	10,71	2,43	3,41
20	1	1	1	0	0	4,23	1,38	10,76	2,4	3,37
21	1	1	0	1	1	-0,11	0,81	1,12	-0,62	0,85
22	1	1	0	1	0	-0,17	0,77	1,18	-0,66	0,79
23	1	1	0	0	1	4,59	2,56	12,24	2,18	5,01
24	1	1	0	0	0	4,54	2,53	12,29	2,15	4,96
25	1	0	1	1	1	-0,19	0	-2,35	-0,34	-0,38
26	1	0	1	1	0	-0,25	-0,04	-2,29	-0,38	-0,43
27	1	0	1	0	1	4,46	1,6	8,78	2,42	3,63
28	1	0	1	0	0	4,4	1,57	8,86	2,39	3,59
29	1	0	0	1	1	0	1,34	-3,58	-0,85	2,37
30	1	0	0	1	0	-0,06	1,29	-3,53	-0,89	2,31
31	1	0	0	0	1	4,68	3,16	7,45	1,86	6,56
32	1	0	0	0	0	4,63	3,12	7,52	1,83	6,51

Tabla III-B: Rendimiento promedio de codificación [% BD-Rate] para la Clase B

#	SCL	SAO	DB	RDOQ	TrSk	SSIM	MS-SSIM	VMAF	VIF	PSNR HVS-M
1	0	1	1	1	1	0	0	0	0	0
2	0	1	1	1	0	-0,07	-0,06	-0,03	-0,06	-0,08
3	0	1	1	0	1	5,27	3,74	12,66	5,59	7,18
4	0	1	1	0	0	5,22	3,7	12,69	5,56	7,13
5	0	1	0	1	1	0,33	1,4	1,24	0,08	2,12
6	0	1	0	1	0	0,25	1,34	1,22	0,01	2,03
7	0	1	0	0	1	5,67	5,35	14,06	5,71	9,52
8	0	1	0	0	0	5,62	5,31	14,08	5,68	9,47
9	0	0	1	1	1	0,4	0,49	-1,92	-0,06	0,45
10	0	0	1	1	0	0,32	0,43	-1,95	-0,13	0,37
11	0	0	1	0	1	5,68	4,22	10,34	5,36	7,62
12	0	0	1	0	0	5,63	4,18	10,34	5,33	7,56
13	0	0	0	1	1	0,48	2,3	-2,94	-0,2	4,12
14	0	0	0	1	0	0,4	2,23	-2,98	-0,27	4,03
15	0	0	0	0	1	5,78	6,38	9,23	5,23	11,68
16	0	0	0	0	0	5,73	6,34	9,23	5,19	11,62
17	1	1	1	1	1	-0,72	-0,48	-0,47	-0,6	-1,06
18	1	1	1	1	0	-0,78	-0,52	-0,52	-0,71	-1,17
19	1	1	1	0	1	2,76	1,91	9,85	2,53	3,14
20	1	1	1	0	0	2,71	1,87	9,85	2,45	3,06
21	1	1	0	1	1	-0,4	0,9	0,75	-0,57	0,95
22	1	1	0	1	0	-0,45	0,85	0,71	-0,68	0,83
23	1	1	0	0	1	3,16	3,45	11,26	2,58	5,22
24	1	1	0	0	0	3,11	3,41	11,25	2,51	5,13
25	1	0	1	1	1	-0,34	-0,01	-2,47	-0,7	-0,63
26	1	0	1	1	0	-0,39	-0,05	-2,52	-0,82	-0,74
27	1	0	1	0	1	3,17	2,37	7,59	2,33	3,56
28	1	0	1	0	0	3,11	2,34	7,57	2,24	3,46
29	1	0	0	1	1	-0,26	1,77	-3,5	-0,89	2,92
30	1	0	0	1	0	-0,31	1,73	-3,55	-1,01	2,8
31	1	0	0	0	1	3,27	4,46	6,5	2,11	7,32
32	1	0	0	0	0	3,22	4,41	6,47	2,02	7,22

Tabla III-C: Rendimiento promedio de codificación [% BD-Rate] para la Clase C

#	SCL	SAO	DB	RDOQ	TrSk	SSIM	MS-SSIM	VMAF	VIF	PSNR HVS-M
1	0	1	1	1	1	0	0	0	0	0
2	0	1	1	1	0	-0,23	-0,22	-0,16	-0,18	-0,26
3	0	1	1	0	1	2,1	2,89	10,43	4,54	5
4	0	1	1	0	0	1,98	2,76	10,34	4,48	4,81
5	0	1	0	1	1	0,79	1	0,85	0,38	1,65
6	0	1	0	1	0	0,57	0,78	0,67	0,2	1,38
7	0	1	0	0	1	3,01	4,06	11,25	4,95	6,75
8	0	1	0	0	0	2,89	3,92	11,14	4,88	6,55
9	0	0	1	1	1	0,71	0,59	-1,84	0,25	0,65
10	0	0	1	1	0	0,5	0,38	-2,06	0,07	0,4
11	0	0	1	0	1	2,89	3,56	8,02	4,78	5,76
12	0	0	1	0	0	2,79	3,44	7,85	4,72	5,58
13	0	0	0	1	1	1,39	2,02	-2,71	0,72	3,77
14	0	0	0	1	0	1,17	1,8	-2,93	0,54	3,51
15	0	0	0	0	1	3,73	5,33	7,13	5,32	9,14
16	0	0	0	0	0	3,63	5,19	6,96	5,25	8,95
17	1	1	1	1	1	-0,02	-0,2	-0,32	-0,09	-0,22
18	1	1	1	1	0	-0,27	-0,44	-0,55	-0,34	-0,56
19	1	1	1	0	1	0,8	1,22	7,75	2,99	2,97
20	1	1	1	0	0	0,59	0,99	7,46	2,83	2,66
21	1	1	0	1	1	0,77	0,75	0,51	0,25	1,35
22	1	1	0	1	0	0,51	0,5	0,26	-0,01	0,99
23	1	1	0	0	1	1,68	2,28	8,59	3,34	4,59
24	1	1	0	0	0	1,48	2,05	8,31	3,17	4,27
25	1	0	1	1	1	0,67	0,36	-2,24	0,14	0,37
26	1	0	1	1	0	0,42	0,13	-2,53	-0,12	0,04
27	1	0	1	0	1	1,57	1,85	5,4	3,2	3,65
28	1	0	1	0	0	1,37	1,62	5,02	3,04	3,35
29	1	0	0	1	1	1,32	1,74	-3,12	0,56	3,44
30	1	0	0	1	0	1,08	1,48	-3,41	0,29	3,08
31	1	0	0	0	1	2,36	3,5	4,49	3,66	6,94
32	1	0	0	0	0	2,16	3,26	4,12	3,5	6,63

Tabla III-D: Rendimiento promedio de codificación [% BD-Rate] para la Clase D

#	SCL	SAO	DB	RDOQ	TrSk	SSIM	MS-SSIM	VMAF	VIF	PSNR HVS-M
1	0	1	1	1	1	0	0	0	0	0
2	0	1	1	1	0	-0,12	-0,21	-0,22	-0,2	-0,28
3	0	1	1	0	1	-0,42	0,97	10,54	4,34	4,77
4	0	1	1	0	0	-0,45	0,78	10,4	4,25	4,55
5	0	1	0	1	1	1,76	0,38	0,6	0,06	0,92
6	0	1	0	1	0	1,61	0,15	0,33	-0,15	0,63
7	0	1	0	0	1	1,56	1,48	11,05	4,44	5,8
8	0	1	0	0	0	1,51	1,27	10,91	4,35	5,55
9	0	0	1	1	1	0,41	0,1	-1,94	0,09	0,28
10	0	0	1	1	0	0,29	-0,12	-2,27	-0,1	0
11	0	0	1	0	1	0,04	1,12	8,25	4,45	5,14
12	0	0	1	0	0	0,04	0,94	7,97	4,36	4,9
13	0	0	0	1	1	3,19	0,71	-2,68	0,22	2,27
14	0	0	0	1	0	3,05	0,47	-3	0,02	1,97
15	0	0	0	0	1	3,25	1,99	7,49	4,63	7,32
16	0	0	0	0	0	3,25	1,8	7,22	4,54	7,08
17	1	1	1	1	1	0,87	0,01	-0,33	-0,03	-0,19
18	1	1	1	1	0	0,71	-0,27	-0,54	-0,28	-0,52
19	1	1	1	0	1	0,67	-0,12	8,07	2,95	2,97
20	1	1	1	0	0	0,67	-0,3	7,82	2,82	2,69
21	1	1	0	1	1	2,56	0,34	0,24	-0,01	0,64
22	1	1	0	1	0	2,39	0,06	0,01	-0,27	0,29
23	1	1	0	0	1	2,53	0,31	8,62	2,98	3,84
24	1	1	0	0	0	2,5	0,12	8,37	2,85	3,54
25	1	0	1	1	1	1,28	0,07	-2,38	0,04	0,06
26	1	0	1	1	0	1,13	-0,21	-2,7	-0,21	-0,27
27	1	0	1	0	1	1,1	-0,01	5,8	3,03	3,28
28	1	0	1	0	0	1,13	-0,17	5,43	2,91	3,01
29	1	0	0	1	1	4,02	0,63	-3,11	0,12	1,97
30	1	0	0	1	0	3,86	0,35	-3,43	-0,14	1,62
31	1	0	0	0	1	4,19	0,76	5,05	3,13	5,34
32	1	0	0	0	0	4,21	0,58	4,67	3,01	5,05

Tabla III-E: Rendimiento promedio de codificación [% BD-Rate] para la Clase E

#	SCL	SAO	DB	RDOQ	TrSk	SSIM	MS-SSIM	VMAF	VIF	PSNR HVS-M
1	0	1	1	1	1	0	0	0	0	0
2	0	1	1	1	0	-0,06	-0,06	-0,08	-0,07	-0,1
3	0	1	1	0	1	2,98	2,26	8,85	3,46	4,65
4	0	1	1	0	0	2,89	2,19	8,72	3,41	4,56
5	0	1	0	1	1	2,14	2,24	1,28	0,42	2,7
6	0	1	0	1	0	2,08	2,17	1,2	0,35	2,59
7	0	1	0	0	1	5,39	4,76	10,31	3,86	7,46
8	0	1	0	0	0	5,3	4,69	10,18	3,81	7,37
9	0	0	1	1	1	0,59	0,74	-2,14	0,28	0,68
10	0	0	1	1	0	0,53	0,67	-2,23	0,21	0,58
11	0	0	1	0	1	3,64	3,1	6,4	3,69	5,31
12	0	0	1	0	0	3,54	3,02	6,29	3,63	5,22
13	0	0	0	1	1	3,35	3,8	-2,61	0,75	4,96
14	0	0	0	1	0	3,29	3,73	-2,69	0,67	4,85
15	0	0	0	0	1	6,77	6,57	5,94	4,15	9,87
16	0	0	0	0	0	6,67	6,49	5,84	4,09	9,77
17	1	1	1	1	1	-0,59	-0,44	-0,25	-0,43	-0,77
18	1	1	1	1	0	-0,64	-0,49	-0,34	-0,5	-0,85
19	1	1	1	0	1	1,5	1,16	7,5	2,25	2,77
20	1	1	1	0	0	1,43	1,11	7,42	2,21	2,71
21	1	1	0	1	1	1,53	1,77	1	-0,03	1,86
22	1	1	0	1	0	1,48	1,72	0,91	-0,11	1,77
23	1	1	0	0	1	3,86	3,6	8,9	2,6	5,47
24	1	1	0	0	0	3,78	3,55	8,82	2,55	5,39
25	1	0	1	1	1	0	0,3	-2,53	-0,18	-0,11
26	1	0	1	1	0	-0,05	0,25	-2,63	-0,25	-0,2
27	1	0	1	0	1	2,16	1,99	4,95	2,46	3,41
28	1	0	1	0	0	2,09	1,95	4,89	2,42	3,34
29	1	0	0	1	1	2,73	3,33	-2,99	0,26	4,1
30	1	0	0	1	0	2,68	3,27	-3,09	0,18	4
31	1	0	0	0	1	5,22	5,38	4,49	2,86	7,82
32	1	0	0	0	0	5,14	5,32	4,42	2,81	7,74

Tabla III-F: Rendimiento promedio de codificación [% BD-Rate] para la Clase F

#	SCL	SAO	DB	RDOQ	TrSk	SSIM	MS-SSIM	VMAF	VIF	PSNR HVS-M
1	0	1	1	1	1	0	0	0	0	0
2	0	1	1	1	0	3,9	4,68	3,38	5,52	4,29
3	0	1	1	0	1	2,33	1,11	5,39	2,2	2,58
4	0	1	1	0	0	6,65	6,21	10,24	9,23	7,43
5	0	1	0	1	1	1,11	1,3	0,32	-0,48	1,55
6	0	1	0	1	0	5,13	6,11	3,73	5,8	5,9
7	0	1	0	0	1	3,53	2,56	5,74	2,46	4,17
8	0	1	0	0	0	7,91	7,74	10,6	9,5	9,09
9	0	0	1	1	1	1,36	1,56	0,7	0,95	1,72
10	0	0	1	1	0	5,43	6,59	3,15	7,95	5,8
11	0	0	1	0	1	3,7	2,77	6,43	3,93	4,29
12	0	0	1	0	0	8,08	8,06	10,04	11,67	8,86
13	0	0	0	1	1	2,56	3,19	0,27	1,26	4,07
14	0	0	0	1	0	6,77	8,38	2,75	8,29	8,31
15	0	0	0	0	1	5,02	4,61	5,99	4,25	6,76
16	0	0	0	0	0	9,52	10,06	9,64	12,02	11,51
17	1	1	1	1	1	-0,29	-0,11	-0,25	-0,82	-0,2
18	1	1	1	1	0	3,55	4,63	3,2	5,39	4,06
19	1	1	1	0	1	1,06	0,37	4,18	1,35	1,45
20	1	1	1	0	0	5,4	5,54	8,99	8,38	6,26
21	1	1	0	1	1	0,8	1,18	0,05	-0,58	1,31
22	1	1	0	1	0	4,76	6,03	3,55	5,64	5,63
23	1	1	0	0	1	2,25	1,8	4,54	1,59	2,98
24	1	1	0	0	0	6,66	7,05	9,45	8,64	7,86
25	1	0	1	1	1	1,04	1,46	0,42	0,84	1,48
26	1	0	1	1	0	5,07	6,55	2,89	7,77	5,54
27	1	0	1	0	1	2,39	1,96	5,2	3,03	3,09
28	1	0	1	0	0	6,82	7,44	8,81	10,76	7,61
29	1	0	0	1	1	2,24	3,05	0	1,12	3,78
30	1	0	0	1	0	6,41	8,3	2,49	8,08	8,01
31	1	0	0	0	1	3,68	3,77	4,77	3,32	5,5
32	1	0	0	0	0	8,24	9,42	8,42	11,09	10,21

métricas VIF y VMAF reportan ganancias y pérdidas según la clase de secuencia, siendo estas cercanas al cero. Se puede decir que, para estas métricas, el efecto de bloques en las secuencias no se tiene en gran consideración a la hora de dar la puntuación final.

Con respecto al incremento de tiempos de CPU, deshabilitar este filtro obtiene de media un ahorro de un 0,3 %, como puede verse en la Tabla IV.

C. Filtro SAO

El filtro SAO, que por defecto se encuentra habilitado, trata de minimizar las distorsiones introducidas por la etapa de cuantización.

En general, observamos un aumento de BD-Rate cuando se deshabilita el filtro SAO, siendo este aumento más importante en la clase F (secuencias de vídeo sintéticas). Sin embargo, la métrica VMAF es la única que evalúa de forma positiva deshabilitar este filtro. Para esta métrica, deshabilitar el filtro SAO manteniendo el parámetro RDOQ habilitado reporta ganancias reales de BD-Rate de hasta un $-3,5\%$ para las clases A a E. La clase F es la única donde la VMAF no reporta ganancias.

La métrica VIF muestra un comportamiento similar al dado por la VMAF, pero únicamente para las clases A y B, siendo sus ganancias nunca superiores a $0,5\%$ de BD-Rate.

Por último, al igual que el filtro de deblocking, deshabilitar el filtro SAO no influye de manera significativa en la complejidad de la codificación, obteniendo en promedio una reducción de tiempos de CPU de un $0,5\%$, tal y como se observa en la Tabla IV.

D. Rate-distortion optimized quantization (RDOQ)

El algoritmo RDOQ consigue estimar un valor de cuantización óptimo que minimiza el coste del compromiso entre la tasa de bits y la distorsión, midiendo esta distorsión mediante la métrica objetiva PSNR. Por defecto se encuentra habilitado, por lo que hemos analizado el impacto perceptual al deshabilitar tanto el parámetro RDOQ como el RDOQTS.

En vista de los resultados obtenidos se puede decir que deshabilitar este algoritmo impli-

ca un aumento de los valores BD-Rate para la mayoría de las clases y métricas perceptuales. Los beneficios de habilitar el algoritmo RDOQ son mayores para la métrica VMAF, donde en promedio obtiene un ahorro de un $9,8\%$ de BD-Rate.

En cuanto a la complejidad del algoritmo, deshabilitar los parámetros RDOQ y RDOQTS implican una reducción media de un $6,82\%$ de tiempos de CPU, alcanzando para $QP = 22$ una reducción de un $15,78\%$, como se ve en la Tabla IV.

E. Transform Skip

Los parámetros TransformSkip y TransformSkipFast se encuentran habilitados por defecto, ya que, para la métrica PSNR, este algoritmo obtiene un importante ahorro de BD-Rate para las secuencias de clase F, mientras que para las demás clases apenas hay cambios.

Analizando el comportamiento de las métricas perceptuales, podemos ver como para las secuencias de clase F (Tabla III-F), deshabilitar los parámetros de TransformSkip resulta en un importante empeoramiento (incremento) de valores BD-Rate, llegando en algunos casos a obtener un 7% de ganancia. Para el resto de las clases, a diferencia de lo que ocurre con la PSNR, la mayoría de las métricas perceptuales reportan ganancias, aunque en ningún caso superiores a $0,4\%$.

En cuanto a las métricas perceptuales, viendo la Tabla III-F, correspondiente a las secuencias de clase F, podemos observar que deshabilitar los parámetros de Transform Skip provoca un importante deterioro (incremento) de los valores de BD-Rate.

Con respecto a la complejidad del algoritmo, de media deshabilitar los parámetros de Transform Skip reduce en un 15% los tiempos de codificación, con independencia de la QP escogida, tal y como puede verse en la Tabla IV.

V. DISCUSIÓN

A raíz de los resultados de la Sección IV, existen herramientas de codificación cuyo comportamiento es evaluado de igual forma por todas las métricas perceptuales, mientras que hay otras donde existen discrepancias en función de

Tabla IV: Aumento/disminución promedio de tiempo de CPU [%] al habilitar/deshabilitar cada herramienta de codificación con respecto a su valor predeterminado (valores negativo corresponden a ahorro de tiempo)

	QP 22	QP 27	QP 32	QP 37	QP 42	Prom.
SCL on	7,41	8,44	6,57	5,18	3,47	6,21
SAO off	-0,3	-0,21	-0,53	-0,37	-0,58	-0,4
DB off	-0,22	-0,18	-0,41	-0,24	-0,26	-0,26
RDOQ off	-16,56	-10,77	-6,4	-1,89	1,5	-6,82
TrSk off	-15,78	-14,83	-14,24	-13,29	-13,22	-14,27

la métrica o de la clase de secuencia. En esta sección vamos a discutir sobre esto, y analizaremos más en detalle los resultados vistos en la sección anterior.

Como se ha visto en la Sección IV-E, al habilitar los parámetros relativos al Transform Skip para la clase F las métricas perceptuales obtienen un gran rendimiento, mientras que para el resto de clases no se obtienen ganancias, sino pérdidas (aunque con valores cercanos a cero). Si a esto sumamos el coste computacional de más de un 15 % al utilizar este algoritmo, podemos establecer que, para las secuencias naturales (clases A a E), no se recomienda el uso de este algoritmo, mientras que para secuencias sintéticas o artificiales (clase F) es altamente recomendable su uso.

En cuanto al parámetro ScalingList, como hemos visto en la Sección IV-A, su uso resulta beneficioso para prácticamente cualquier configuración, obteniendo reducciones de BD-Rate de hasta un 3,8 % (para la VMAF). Este comportamiento era el esperado, ya que esta herramienta hace uso de una cuantización no uniforme basada en la CSF, mejorando la calidad perceptual del vídeo [15]. Por tanto, se recomienda habilitar siempre este parámetro que, por defecto, permanece deshabilitado.

En general, todas las métricas perceptuales reportan incrementos significativos (pérdidas) de BD-Rate al deshabilitar los parámetros RDOQ y RDOQTS. De media, el aumento es de un 4,16 %, alcanzando un 11,5 % para la métrica VMAF en secuencias de vídeo de clase A. Existen algunas excepciones donde deshabilitar este algoritmo resulta beneficioso. Para las métricas SSIM y MS-SSIM, al deshabilitar los parámetros de RDOQ, manteniendo los filtros In-loop habilitados (DB y SAO), se consiguen ganancias de BD-Rate para las secuencias de

clase D (baja resolución). Estas ganancias, sin embargo, no son nada elevadas, alcanzando en el mejor de los casos el 0,4 %. Por tanto, se recomienda no deshabilitar esta herramienta de codificación en ningún caso.

Hasta ahora, hemos visto que, para las métricas perceptuales, ha habido cierta unanimidad en cuanto a los parámetros analizados: para obtener un mejor rendimiento desde el punto de vista perceptual, se deben habilitar los parámetros SCL y RDOQ, mientras que el parámetro TransformSkip debe habilitarse únicamente para secuencias de clase F. Asumiendo esto, vamos a pasar a analizar el comportamiento de los filtros In-loop, que es donde hay mayor discrepancia en cuanto a resultados.

El comportamiento general de las métricas perceptuales en cuanto al uso de los filtros In-loop tiene dos posiciones contrarias. Por un lado, las métricas SSIM, MS-SSIM y PSNR-HVS-M proporcionan mejores resultados cuando los filtros de deblocking y SAO se encuentran habilitados. Por otro lado, se encuentran las métricas VIF y VMAF, que reportan beneficios significativos al deshabilitar ambos filtros, llegando a obtener para el caso de esta última métrica unas ganancias de hasta un 3,4 %. Este comportamiento tiene algunas excepciones, por ejemplo, para las secuencias de clase F, las métricas VIF y VMAF no evalúan de forma positiva el deshabilitar los filtros In-loop. La VIF, para las secuencias de clase C, D y E, tampoco reporta beneficios al deshabilitar los filtros de deblocking y SAO.

Con respecto a la métrica VMAF, se ha comprobado que, de los filtros In-loop, deshabilitar el filtro SAO es el que más afecta en cuanto a reducción de BD-Rate. Desde el grupo de investigación GATCOM hemos reportado este comportamiento de la métrica VMAF a sus desarrolladores, y en su respuesta nos han indicado que van a estudiarlo en profundidad para determinar si se trata de un problema de entrenamiento de su red neuronal o si por el contrario es un comportamiento esperado y correcto.

Aunque todas las métricas de calidad perceptual tratan de caracterizar el sistema visual humano, cada una utiliza una aproximación di-

ferente. Algunas realizan una descomposición en subbandas inspirada en complejos modelos del HVS, otras extraen la información estructural de las escenas, y otras utilizan patrones estadísticos espaciotemporales. Por lo tanto, como se puede ver en los resultados de este estudio, en algunos casos no hay unanimidad entre las métricas perceptuales. En los casos donde las métricas reportan variaciones de BD-Rate en la misma dirección, la conclusión es directa, pero cuando los reportes indican discrepancias en los resultados, se sugiere realizar pruebas subjetivas de tipo MOS para validar los resultados.

Por último, otra métrica de rendimiento que podemos utilizar para evaluar la configuración más adecuada de las herramientas de codificación es su contribución a la complejidad global de la codificación. En la Tabla IV se muestra el porcentaje de reducción (o aumento) de tiempos de CPU al modificar el estado por defecto de cada herramienta de codificación en estudio, mostrando su impacto en la complejidad global en la codificación del estándar HEVC. Si habilitamos todas las herramientas de codificación en estudio excepto el Transform Skip (la mejor configuración perceptual en cuanto a R/D), la complejidad global del estándar HEVC se reducirá en un 14,27% de media, en comparación con la configuración por defecto.

VI. CONCLUSIONES

En este artículo hemos analizado el impacto perceptual de los parámetros de codificación del estándar HEVC. Para ello, hemos codificado un dataset completo de secuencias de vídeo variando los parámetros de codificación del estándar y obtenido los valores BD-Rate para un conjunto de métricas de calidad perceptuales, lo cual nos permite determinar la mejora o empeoramiento con respecto a la configuración predeterminada. Finalmente, hemos analizado el impacto perceptual que tiene cada herramienta de codificación y su relación con el resto de herramientas.

Tras analizar los resultados proporcionados por el conjunto de herramientas de codificación bajo estudio, hemos llegado a las siguientes conclusiones: (a) las herramientas de co-

dificación con mayor rendimiento perceptual son ScalingList y el algoritmo RDOQ, (b) el parámetro TransformSkip obtiene también muy buen rendimiento perceptual pero únicamente para secuencias de clase F, y (c) los filtros In-loop (DB y SAO) tienen comportamientos opuestos dependiendo de la métrica utilizada y la clase de secuencia a codificar.

La configuración recomendada que maximiza el R/D perceptual sería habilitar el algoritmo ScalingList y RDOQ, y deshabilitar el TransformSkip (a no ser que se codifiquen secuencias artificiales o sintéticas, en cuyo caso es recomendable dejarlo habilitado). Con respecto a los filtros In-loop, al no haber consenso entre las diferentes métricas, para determinar la mejor opción se necesitaría llevar a cabo pruebas subjetivas para decidir qué hacer al respecto.

Con todo esto, nuestra intención a la hora de elaborar este artículo es ayudar a otros investigadores a determinar la mejor configuración de codificación posible, en función del tipo de secuencia, que maximice el rendimiento R/D perceptual, teniendo en cuenta también la complejidad de las diferentes herramientas de codificación dadas por el estándar. Además, también pueden ser útil para determinar la métrica perceptual a utilizar en futuros estudios.

Como trabajo futuro, planeamos ampliar este estudio a otros estándares de codificación de vídeo actuales, como AV1 y VVC, incluir más herramientas de codificación y métricas perceptuales, así como realizar pruebas subjetivas para determinar las configuraciones que maximicen el rendimiento R/D perceptual en los casos donde las métricas objetivas no obtengan resultados concluyentes.

AGRADECIMIENTOS

El presente trabajo ha sido financiado por el Ministerio de Ciencia, Innovación y Universidades con referencia RTI2018-098156-B-C54 cofinanciado con fondos FEDER (MINECO/-FEDER/UE) y por la Conselleria de Innovación, Universidades, Ciencia y Sociedad Digital de la Comunidad Valenciana con referencia GV/2019/020.

REFERENCIAS

- [1] H. R. Wu and K. R. Rao, *Digital Video Image Quality and Perceptual Coding (Signal Processing and Communications)*, CRC Press, Inc., USA, 2005.
- [2] Y. Zhang, M. Naccari, D. Agrafiotis, M. Mrak, and D. R. Bull, "High dynamic range video compression by intensity dependent spatial quantization in HEVC," *2013 Picture Coding Symposium (PCS)*, pp. 353–356, 2013.
- [3] Y. Zhang, M. Naccari, D. Agrafiotis, M. Mrak, and D. R. Bull, "High Dynamic Range Video Compression Exploiting Luminance Masking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 5, pp. 950–964, 2016.
- [4] B. Girod, *What's Wrong with Mean-Squared Error?*, p. 207–220, MIT Press, Cambridge, MA, USA, 1993.
- [5] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Transactions on Communications*, vol. 43, no. 12, pp. 2959–2965, 1995.
- [6] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thirty-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, Nov 2003, vol. 2, pp. 1398–1402 Vol.2.
- [7] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr 2004.
- [8] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electronics Letters*, vol. 44, no. 13, pp. 800–801, Jun 2008.
- [9] Zhi Li, Anne Aaron, Ioannis Katsavounidis, Anush Moorthy, and Megha Manohara, "Toward a Practical Perceptual Video Quality Metric," Jun 2016.
- [10] Miguel Onofre Martínez-Rach, *Perceptual image coding for wavelet based encoders*, Ph.D. thesis, Universidad Miguel Hernández de Elche, Dec 2014.
- [11] J. Mannos and D. Sakrison, "The effects of a visual fidelity criterion of the encoding of images," *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 525–536, Jul 1974.
- [12] N. Nill, "A Visual Model Weighted Cosine Transform for Image Compression and Quality Assessment," *IEEE Trans. on Communications*, vol. 33, no. 6, pp. 551–557, Jun 1985.
- [13] B. Chitprasert and K. R. Rao, "Human visual weighted progressive image transmission," *IEEE Trans. on Communications*, vol. 38, no. 7, pp. 1040–1044, Jul 1990.
- [14] M. Wien, *High Efficiency Video Coding: Coding Tools and Specifications*, Signals and Communication Technology. Springer, 2015.
- [15] Munsif Haque, Ali Tabatabai, and Yoshitaka Morigami, "HVS model based default quantization matrices," in *7th Meeting of the Joint Collaborative Team on Video Coding (JCT-VC)*. Document: JCTVC-G880, Nov 2011.
- [16] A. Norkin, G. Bjontegaard, A. Fuldseth, M. Narroschke, M. Ikeda, K. Andersson, M. Zhou, and G. Van der Auwera, "HEVC Deblocking Filter," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1746–1754, Dec 2012.
- [17] C. Fu, E. Alshina, A. Alshin, Y. Huang, C. Chen, C. Tsai, C. Hsu, S. Lei, J. Park, and W. Han, "Sample Adaptive Offset in the HEVC Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1755–1764, Dec 2012.
- [18] Gisele Bjontegaard, "Calculation of average PSNR differences between RD-Curves," in *Proc. of the ITU-T Video Coding Experts Group - Thirteenth Meeting*, Apr 2001.
- [19] J. Stankowski, C. Korzeniewski, M. Domański, and T. Grajek, "Rate-distortion optimized quantization in HEVC: Performance limitations," in *2015 Picture Coding Symposium (PCS)*, 2015, pp. 85–89.
- [20] Lee Prangnell, "Visually lossless coding in HEVC: A high bit depth and 4:4:4 capable JND-based perceptual quantisation technique for HEVC," *Signal Processing: Image Communication*, vol. 63, pp. 125–140, 2018.
- [21] Cuiling Lan, Jizheng Xu, Gary J. Sullivan, and Feng Wu, "Intra transform skipping," in *9th Meeting of the Joint Collaborative Team on Video Coding (JCT-VC)*. Document: JCTVC-I0408, Apr 2012.
- [22] Frank Bossen, "Common test conditions and software reference," in *11th Meeting of the Joint Collaborative Team on Video Coding (JCT-VC)*. Doc. JCTVC-K1100, Oct 2012.
- [23] Fraunhofer Institute for Telecommunications, "HM Reference Software Version 16.20," Sep 2018.
- [24] H. R. Sheikh and A. C. Bovik, "A visual information fidelity approach to video quality assessment," in *The First International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2005, pp. 23–25.
- [25] Nikolay Ponomarenko, Flavia Silvestri, Karen Egiazarian, Marco Carli, Jaakko Astola, and Vladimir Lukin, "On between-coefficient contrast masking of DCT basis functions," in *Proceedings of the third international workshop on video processing and quality metrics*, 2007, vol. 4.
- [26] S. Li, F. Zhang, L. Ma, and K. N. Ngan, "Image Quality Assessment by Separately Evaluating Detail Losses and Additive Impairments," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 935–949, 2011.
- [27] Recommendation ITU-T P.910, "Subjective video quality assessment methods for multimedia applications," *International Telecommunication Union*, Apr 2008.
- [28] R. Rassool, "VMAF reproducibility: Validating a perceptual practical video quality metric," in *2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Jun 2017, pp. 1–2.
- [29] J. Gutiérrez, L. Krasula, P. Le Callet, Z. Li, and I. Katsavounidis, "VMAF framework performance on UHD videos," in *VQEG Meeting 2017*, Los Gatos, CA, USA, May 2017.
- [30] C. Lee, S. Woo, S. Baek, J. Han, J. Chae, and J. Rim, "Comparison of objective quality models for adaptive bit-streaming services," in *2017 8th*

International Conference on Information, Intelligence, Systems Applications (IISA), 2017, pp. 1–4.

- [31] Universidad Miguel Hernández de Elche, “Grupo de Arquitectura y Tecnología de Computadores,” http://atc.umh.es/gatcom/Ficheros/overall_results.zip.