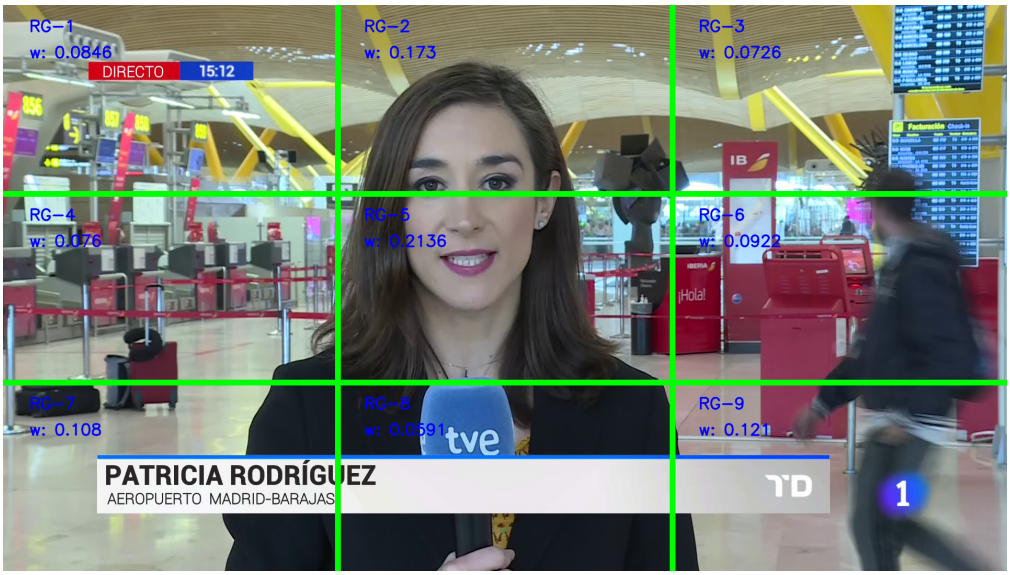


**Figura 4.9:** Detección de saliencia en una secuencia de prueba de La1 HD.

única región. Del mismo modo, en la Figura C.14 se utilizan dos regiones, tres regiones en la Figura C.15, cuatro regiones en la Figura C.16, cinco regiones en la Figura C.17, seis regiones en la Figura C.18, siete regiones en la Figura C.19, ocho regiones en la Figura C.20 y nueve regiones en la Figura C.21.

El resultado final del método de detección de saliencia espacio-temporal propuesto para esta secuencia de vídeo se muestra en la Figura 4.9. La Figura 4.9 representa la distribución de pesos sobre la imagen original, reflejando la importancia relativa de cada una de las regiones en función del tipo de información contenida en cada una de ellas.

Para esta medida de vídeo específica, la asignación de pesos revela que las regiones de mayor importancia, en orden descendente, son: región RG-5 (0.21), región RG-2 (0.17), región RG-9 (0.12), región RG-7 (0.11), región RG-6 (0.09), región RG-1 (0.08), región RG4 (0.08), región RG-3 (0.07) y región RG-8 (0.06). La distribución de estos valores pone de manifiesto la influencia de la detección de rostros en la determinación de la relevancia de cada región, ya que la presencia de un rostro en la imagen, localizado en la región RG-2 y región RG-5, ha resultado en una mayor ponderación de éstas en comparación con el resto de regiones. La presencia de los elementos gráficos en la región RG-7 y región RG-9 también se manifiesta en pesos de ponderación elevados para estas dos regiones en particular.

El proceso de obtención del mapa de saliencia espacio-temporal implica la consideración de diversas alternativas metodológicas, siendo una de las principales decisiones la elección del método de saliencia estática. Las pruebas realizadas han evidenciado un mejor desempeño del método *Spectral Residual* frente al método *Fine Grained*, en términos de menor error de valor de MOS utilizando el enfoque por combinación de regiones basado en saliencia. La Figura 4.10 presenta una comparativa entre ambos métodos con todo el conjunto de datos de prueba, mostrando los resultados del error de MOS en función del número de regiones utilizadas en el análisis.

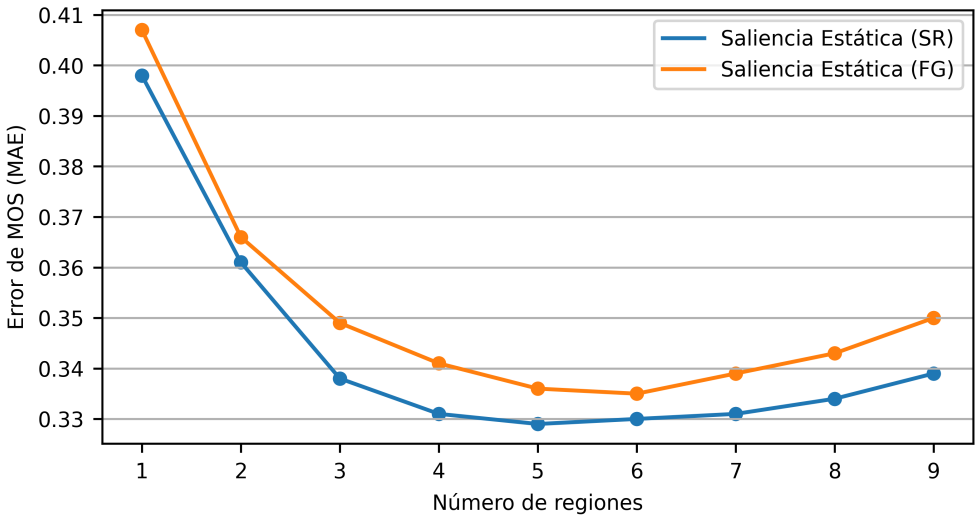


**Figura 4.10:** Error de MOS en el enfoque por combinación de regiones basado en saliencia (1).

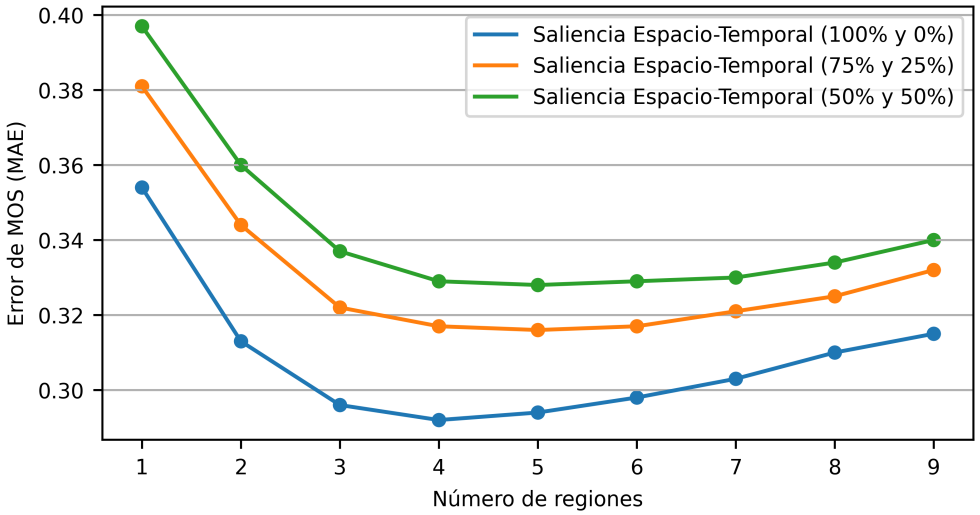


**Figura 4.11:** Error de MOS en el enfoque por combinación de regiones basado en saliencia (2).

En lo que respecta a la fusión entre la saliencia espacial y la saliencia temporal, se han evaluado las distintas combinaciones de ponderación propuestas para determinar su impacto en la estimación final de la calidad. Los resultados experimentales indican que la mejor combinación se alcanza cuando la contribución de saliencia temporal es nula en la fusión final ($\beta = 0.0$). Esto sugiere que la información de la saliencia espacial es suficiente para la caracterización de las regiones de interés, sin necesidad de incorporar la componente relacionada con el movimiento. La Figura 4.11 presenta la comparativa con las diferentes combinaciones propuestas.

El enfoque por combinación de regiones basado en saliencia queda configurado de la siguiente manera: método SR para el cálculo de la saliencia estática y pesos de ponderación (100 %, 0 %) para la obtención de la saliencia espacio-temporal. Con estos parámetros, la Tabla 4.10

**Tabla 4.10:** Error de MOS vs. coste computacional en el enfoque por combinación de regiones basado en saliencia.

| Modo | Nº regiones | Resolución | Coste computacional (%) | Error de MOS (MAE) |
|---|---|---|---|---|
| RG-N1-S | 5 | 640x360 (x1) | 11.11 | 0.35 |
| RG-N2-S | 5, 7 | 640x360 (x2) | 22.22 | 0.31 |
| RG-N3-S | 5, 7, 8 | 640x360 (x3) | 33.33 | 0.30 |
| RG-N4-S | 5, 7, 8, 2 | 640x360 (x4) | 44.44 | 0.29 |
| RG-N5-S | 5, 7, 8, 2, 9 | 640x360 (x5) | 55.55 | 0.29 |
| RG-N6-S | 5, 7, 8, 2, 9, 1 | 640x360 (x6) | 66.66 | 0.30 |
| RG-N7-S | 5, 7, 8, 2, 9, 1, 4 | 640x360 (x7) | 77.77 | 0.30 |
| RG-N8-S | 5, 7, 8, 2, 9, 1, 4, 3 | 640x360 (x8) | 88.88 | 0.31 |
| RG-N9-S | 5, 7, 8, 2, 9, 1, 4, 3, 6 | 640x360 (x9) | 100.00 | 0.32 |

recoge los valores de este enfoque en función del número de regiones utilizadas para el análisis, desde el punto del vista del error de MOS y el coste computacional del enfoque.

La Figura 4.12 ilustra los resultados de este enfoque, mostrando la distribución del error de MOS en función del número de regiones consideradas. La Figura 4.12 también incluye los resultados obtenidos con el enfoque por combinación de regiones basado en promedios. El intervalo de error obtenido con el método de saliencia, comprendido entre 0.35 y 0.29, pone de manifiesto la mejoría del método propuesto, frente al enfoque basado en promedios.

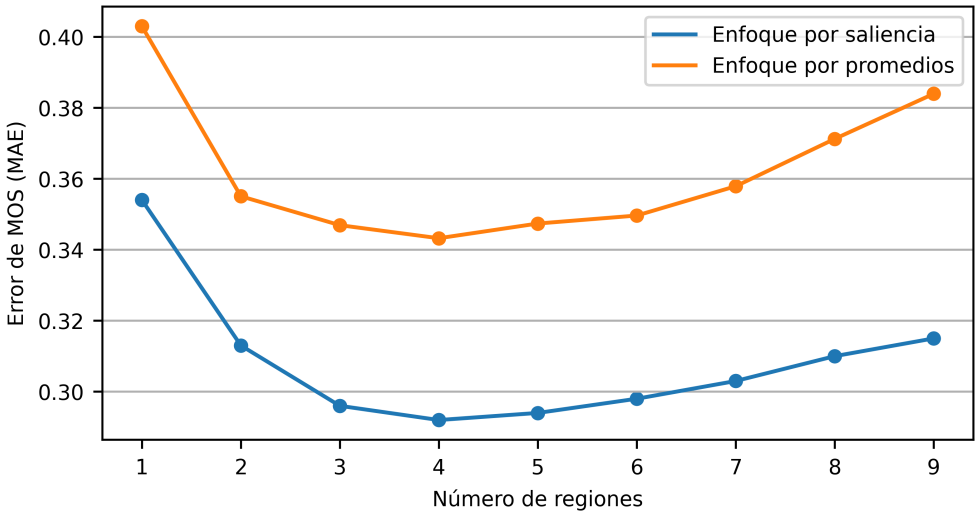


**Figura 4.12:** Error de MOS vs. número de regiones en el enfoque por combinación de regiones basado en saliencia.

La comparativa de ambos enfoques en términos de error de MOS revela una mejor capacidad predictiva al emplear la saliencia en lugar de una simple ponderación con valores promedio

**Tabla 4.11:** Error en la extracción de características en el enfoque por combinación de regiones basado en saliencia.

| Métrica de evaluación | Error (MAE, en %) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | RG-N1-S | RG-N2-S | RG-N3-S | RG-N4-S | RG-N5-S | RG-N6-S | RG-N7-S | RG-N8-S | RG-N9-S |
| Spatial Information | 17.18 | 12.71 | 10.01 | 7.68 | 6.20 | 5.08 | 4.21 | 3.71 | 3.62 |
| Temporal Information | 11.42 | 8.65 | 7.15 | 6.06 | 5.16 | 4.49 | 4.06 | 3.80 | 3.72 |
| Blurring | 44.32 | 33.98 | 28.13 | 24.21 | 20.50 | 17.71 | 15.36 | 13.23 | 11.98 |
| Brightness | 9.19 | 7.21 | 5.92 | 5.16 | 4.45 | 3.92 | 3.46 | 3.14 | 2.84 |
| Contrast | 14.83 | 13.18 | 12.43 | 12.21 | 12.54 | 12.95 | 13.33 | 13.78 | 14.17 |
| Ringing | 21.07 | 15.21 | 12.19 | 9.70 | 8.48 | 7.72 | 7.07 | 6.75 | 6.65 |
| Blockloss | 4.46 | 3.34 | 2.84 | 2.42 | 2.19 | 1.95 | 1.81 | 1.73 | 1.62 |
| Blocking | 19.46 | 14.66 | 12.25 | 10.63 | 9.31 | 8.21 | 7.15 | 6.29 | 5.80 |
| **Vector de caract.** | **17.74** | **13.62** | **11.37** | **9.76** | **8.61** | **7.75** | **7.06** | **6.55** | **6.30** |

entre regiones. Empleando cuatro regiones para la estimación de calidad, mientras que en el modo RG-N4-P (enfoque basado en promedios) se obtiene un error de MOS de 0.34, con el modo RG-N4-S (enfoque basado en saliencia) el error de MOS se reduce hasta un valor de 0.29.

En términos de error en las métricas de vídeo, los resultados obtenidos evidencian una tendencia clara: a medida que se incrementa el número de regiones utilizadas en la estimación de calidad, tanto el error de cada una de las métricas de vídeo de manera individual como el error asociado al vector de características (definido como valor promedio resultante de los valores de las ocho métricas de vídeo), tiende a disminuir. La Tabla 4.11 recoge estos resultados en detalle.

Al analizar el error en el vector de características, los valores obtenidos oscilan entre 17.74 % y 6.30 %. Estos valores son comparables a los obtenidos en el enfoque por combinación de regiones basado en promedios, donde los errores fluctuaban entre 18.08 % y 4.97 %. Con cuatro regiones, el error en el vector de características es de 8.76 % en el modo RG-N4-P y de 9.76 % en el modo RG-N4-S.

Las pruebas realizadas han mostrado que el tiempo medio requerido para la detección de saliencia por cada secuencia de tres segundos de vídeo es de 25.76 ms para la saliencia espacial con el método SR, 55.53 ms para la saliencia espacial con el método FG, 33.90 ms para la saliencia temporal, 59.77 ms para la saliencia espacio-temporal con el método SR y 90.18 ms para la saliencia espacio-temporal con el método FG. La Figura 4.13 recoge el tiempo empleado por el método propuesto de detección de saliencia para cada secuencia de vídeo. El método de detección de saliencia presentado para este enfoque supone un tiempo medio por secuencia de vídeo de 0.0258 s, al emplear el método SR para calcular la saliencia estática y omitiendo el cálculo de la saliencia temporal.

Aunque el enfoque basado en saliencia presenta un desempeño superior al enfoque basado en promedios, el error en la estimación de MOS sigue siendo elevado. Este aspecto, sumado al tiempo medio asociado al método de detección de saliencia y a la limitación de procesar la

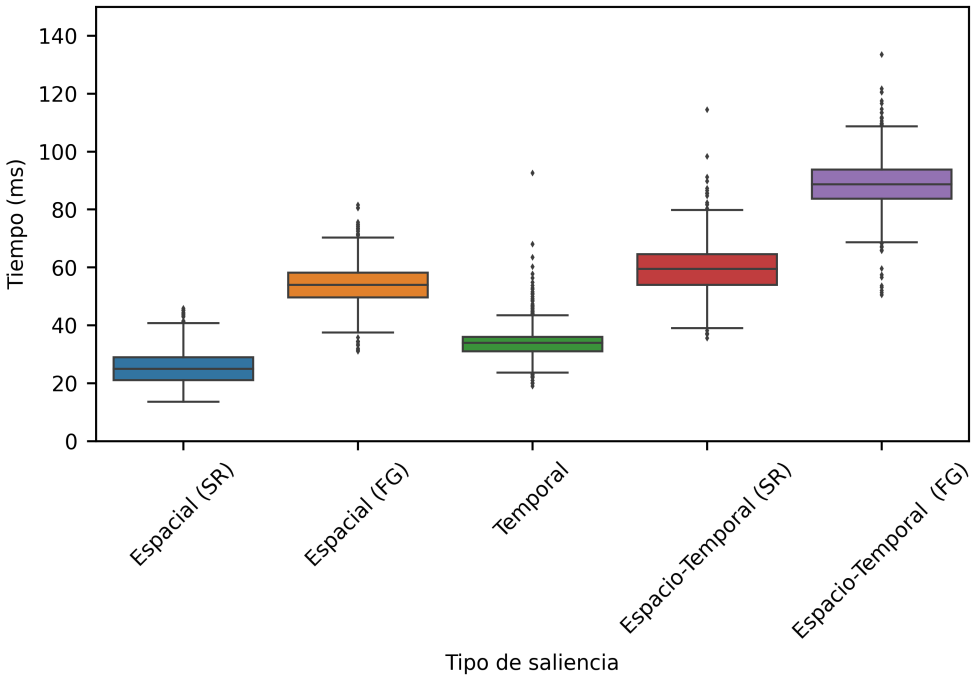


**Figura 4.13:** Tiempo medio de detección de saliencia por secuencia de vídeo.

saliencia únicamente en las dos primeras imágenes de la medida, hace que esta estrategia no sea adecuada para la optimización del coste computacional en la herramienta Video-MOS.

### 4.1.2 Enfoque por cambio de resolución

Un cambio de resolución conlleva un ajuste en el tamaño de la imagen. Para este propósito, la biblioteca de OpenCV proporciona la función *resize*, que permite modificar la resolución de una imagen mediante diferentes métodos de interpolación de píxeles, entre los que se incluyen: *Linear*, *Cubic*, *Area*, *Nearest* y *Lanczos* [8]. En este enfoque de cambio de resolución se ha optado por utilizar el método de interpolación *Cubic*. Esta elección se fundamenta en un análisis comparativo de los distintos métodos de interpolación proporcionados por la biblioteca OpenCV, aplicado a todas las imágenes individuales del conjunto de prueba, y priorizando un equilibrio entre el tiempo requerido para el proceso de cambio de resolución y la calidad obtenida de la nueva imagen reducida.

La calidad de las imágenes resultantes tras el proceso de cambio de resolución se ha evaluado mediante la métrica SSIM, una métrica FR ampliamente utilizada para la comparación entre imágenes de igual tamaño [127]. La métrica SSIM ha demostrado ser efectiva en la evaluación de la calidad de imagen debido a su simplicidad y a su grado de correlación con la percepción visual humana, como se ha evidenciado en diversos estudios comparativas sobre métricas FR [231], [232].

SSIM evalúa la similitud entre una imagen de referencia y su versión distorsionada, midiendo

[8] https://docs.opencv.org/4.11.0/da/d54/group__imgproc__transform.html

**Tabla 4.12:** Comparativa del tipo de interpolación para el cambio de resolución.

| **Método de interpolación** | **Tiempo medio (ms)** | **Valor medio de SSIM** |
|---|---|---|
| *Linear* | 1.52 | 0.92 |
| *Cubic* | 1.62 | 0.93 |
| *Area* | 1.88 | 0.90 |
| *Nearest* | 1.37 | 0.90 |
| *Lanczos* | 2.76 | 0.93 |

diferencias en términos de luminancia ($\mu_x$ y $\mu_y$), contraste ($\sigma_x$ y $\sigma_y$) y estructura ($\sigma_{xy}$). Estas tres componentes convierten a la métrica SSIM en una métrica robusta para evaluar la calidad visual y se basa en el supuesto que la información estructural y las distorsiones estructurales de una imagen tienen una mayor importancia en la percepción visual que las distorsiones no estructurales, como cambios de brillo, cambios de contraste, o cambios de color. Su interpretación es intuitiva, ya que el valor resultado se normaliza en un rango de 0 a 1, donde un valor de SSIM igual a 1 indica una coincidencia perfecta entre ambas imágenes, mientras que valores menores reflejan un mayor nivel de distorsión.

La fórmula matemática de la métrica SSIM queda definida en la Ecuación 4.2.

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{4.2}$$

Donde:

- $x$ e $y$ representan las dos imágenes que se están comparando.
- $\mu_x$ y $\mu_y$ representan las medias de las imágenes $x$ e $y$, respectivamente.
- $\sigma_x^2$ y $\sigma_y^2$ representan las varianzas de las imágenes $x$ e $y$, respectivamente.
- $\sigma_{xy}$ representa la covarianza entre $x$ e $y$.
- $C_1$ y $C_2$ representan constantes. Típicamente, $C_1 = (K_1 L)^2$ y $C_2 = (K_2 L)^2$, donde $L$ es el rango dinámico de los valores de píxeles ($L = 255$ para imágenes de 8 bits), $K_1 = 0{,}01$ y $K_2 = 0{,}03$.

Los datos presentados en la Tabla 4.12, junto con la representación gráfica de la Figura 4.14 y Figura 4.15, muestran los resultados obtenidos para cada método de interpolación de OpenCV. Para el cálculo de la métrica SSIM, se ha empleado un doble proceso de cambio de resolución: primero reduciendo la imagen a 480x270 píxeles y posteriormente volviendo a la resolución original de 1920x1080 píxeles. De este modo, la métrica SSIM se calcula comparando la imagen original con la imagen resultante tras los dos procesos de cambio de resolución. En términos de compromiso entre tiempo de procesamiento y calidad de imagen reconstruida, el método de interpolación *Cubic* ha demostrado ser el método más adecuado.

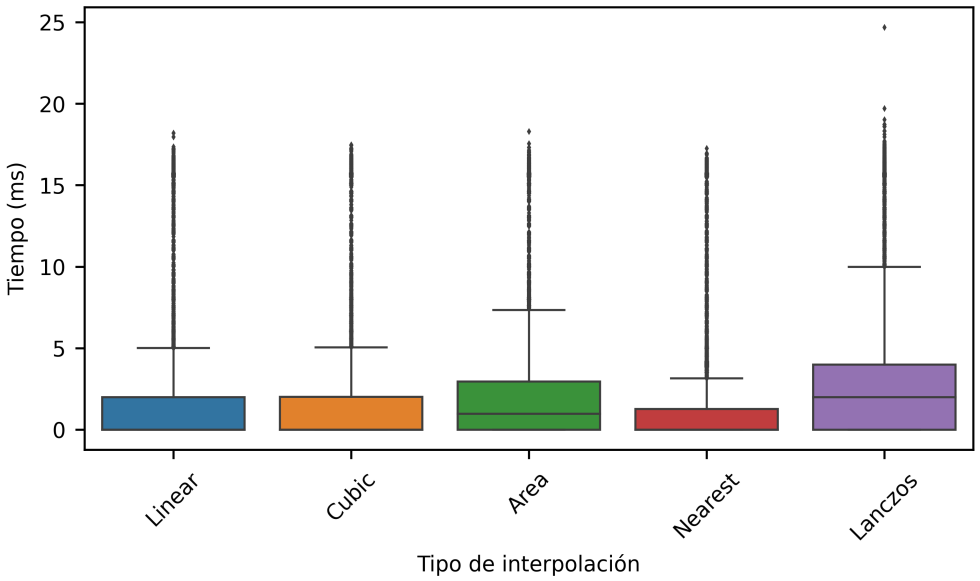


**Figura 4.14:** Tiempo medio de cambio de resolución por imagen según el tipo de interpolación.

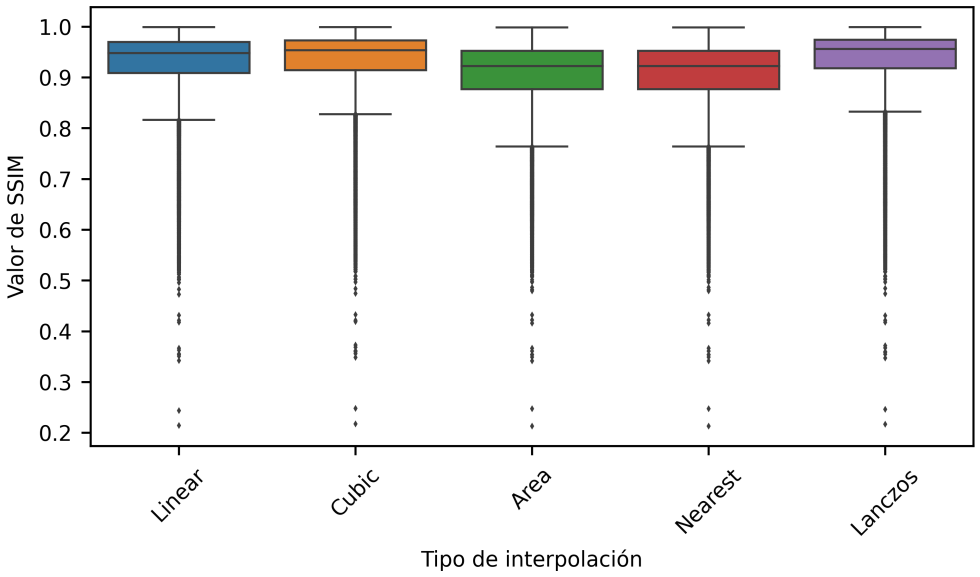


**Figura 4.15:** Valor de SSIM en cambio de resolución por imagen según el tipo de interpolación.

El tiempo medio empleado para hacer un cambio de resolución a 480x270 píxeles utilizando el método de interpolación *Cubic*, sobre el conjunto de imágenes de prueba, es de 1.62 ms por imagen. Este valor resulta prácticamente despreciable en comparación con el tiempo requerido para la extracción de todas las características de la imagen a 1920x1080 píxeles (0.543 s).

El cambio de resolución conlleva un proceso de submuestreo de los píxeles de la imagen. Aunque la información original se conserva en términos de valores de píxeles (proceso de interpolación), dicho submuestreo implica una pérdida de las altas frecuencias, lo que se

**Tabla 4.13:** Error de MOS vs. coste computacional en el enfoque por cambio de resolución.

| Modo | Resolución | Coste computacional (%) | Error de MOS (MAE) |
|---|---|---|---|
| CR-720 | 1280x720 | 44.44 | 0.53 |
| CR-540 | 960x540 | 25 | 0.71 |
| CR-360 | 640x360 | 11..11 | 0.85 |
| CR-270 | 480x270 | 6.25 | 0.93 |
| CR-180 | 320x180 | 2.78 | 0.82 |

traduce en un efecto de desenfoque, pérdida de nitidez y nivel de detalle, alteraciones en la estructura de la imagen y en la información de los bordes. Los resultados obtenidos para el enfoque por cambio de resolución con el método *Cubic* para las diferentes resoluciones propuestas en esta investigación se muestran en la Tabla 4.13. En términos de error de MOS, los valores son muy elevados, oscilando entre 0.53 y 0.93 según el cambio de resolución aplicado.

La Figura 4.16 representa la relación entre el error de MOS y el coste computacional para este enfoque basado en cambio de resolución para las cinco resoluciones propuestas. La tendencia observada en la curva indica que el error en la estimación de calidad disminuye a media que la resolución de la imagen se aproxima al tamaño de la imagen original.

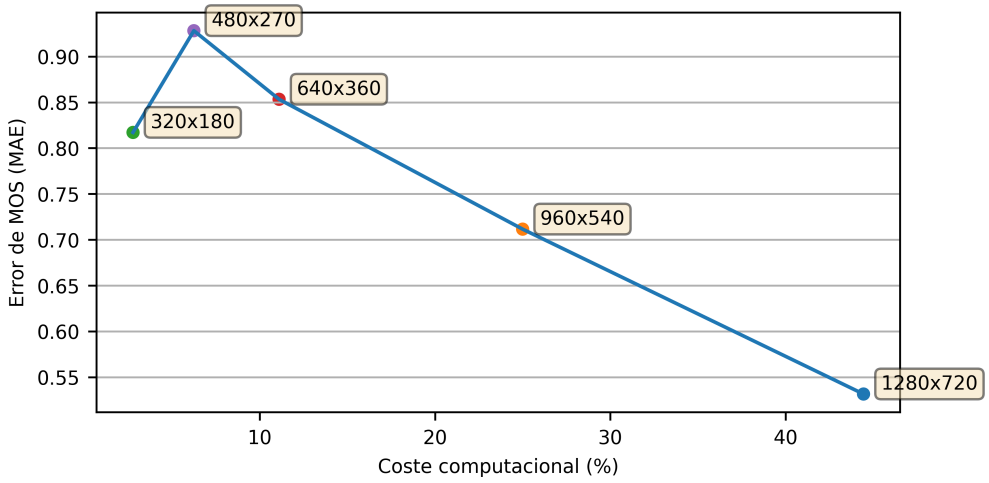


**Figura 4.16:** Error de MOS vs. coste computacional en el enfoque por cambio de resolución.

El análisis de los datos revela diferencias significativas en la información extraída a distintos tamaño de imagen. La Tabla 4.14 recoge los errores obtenidos tanto a nivel de métricas de vídeo individuales como en el cálculo del error promedio sobre el vector de características. En este sentido, el error en el vector de características varía notablemente en función del tamaño de imagen utilizado en el cambio de resolución, alcanzando un valor de 14.42 % para

**Tabla 4.14:** Error en la extracción de características en el enfoque por cambio de resolución.

| **Métrica de evaluación** | **Error (MAE, en %)** | | | | |
|---|---|---|---|---|---|
| | **CR-180** | **CR-270** | **CR-360** | **CR-540** | **CR-720** |
| *Spatial Information* | 47.21 | 39.34 | 32.29 | 20.97 | 11.90 |
| *Temporal Information* | 9.01 | 6.56 | 5.32 | 4.22 | 3.68 |
| *Blurring* | 122.98 | 91.68 | 76.21 | 65.73 | 64.04 |
| *Brightness* | 0.36 | 0.55 | 0.41 | 0.30 | 0.24 |
| *Contrast* | 1.96 | 1.35 | 1.39 | 0.67 | 0.63 |
| *Ringing* | 73.19 | 55.25 | 44.62 | 27.45 | 14.92 |
| *Blockloss* | 12.50 | 20.83 | 16.67 | 4.17 | 4.17 |
| *Blocking* | 33.33 | 30.80 | 28.47 | 21.33 | 15.78 |
| **Vector de caract.** | **37.57** | **30.80** | **25.67** | **18.10** | **14.42** |

la resolución 1280x720 y de 37.57 % para la resolución 320x180.

Las métricas de evaluación de vídeo empleadas en la solución Video-MOS pueden agruparse en dos categorías principales, según recoge la Tabla 3.2: métricas de vídeo basadas a nivel de imagen y métricas basadas a nivel de píxel.

- **Métricas basadas a nivel de imagen**. Este conjunto incluye las métricas *Spatial Information*, *Blurring*, *Ringing*, *Blockloss* y *Blocking*. Todas ellas se fundamentan en la información de bordes y altas frecuencias de la imagen, así como en la aplicación de filtros de procesamiento de imágenes, como el operador de Sobel o el operador Laplaciano. Además, algunas de ellas dependen de la segmentación de la imagen en bloques de 8x8 o de 16x16 píxeles, debido a características propias de la codificación del vídeo y de la transformación de la DCT. La aplicación de estas métricas de vídeo con resoluciones distintas genera variaciones significativas en los resultados, lo que se traduce en errores elevados, tal y como se muestra en la Tabla 4.14. Por ejemplo, al realizar un cambio de resolución a 320x180 píxeles, se obtienen errores de 12.50 % en la métrica de vídeo *Blockloss*, 33.33 % en la métrica *Blocking*, 47.21 % en la métrica de *Spatial Information* y 122.98 % en la métrica de *Blurring*.
- **Métricas basadas a nivel de píxel**. Este conjunto está formado por las métricas de *Temporal Information*, de *Brightness* y de *Contrast*. Estas métricas se calculan directamente a partir de los valores de los píxeles de la imagen, lo que las hace menos sensibles a cambios de resolución, ya que la reducción de tamaño implica un submuestreo que mantiene una representación global de la información original. En el caso de la resolución de 320x180, las métricas basadas en valores de los píxeles presentan errores significativamente inferiores al conjunto de métricas basadas a nivel de la imagen, por debajo de un valor error del 10 %: 0.36 % en la métrica de *Brightness*, 1.96 % en *Contrast* y 9.01 % en *Temporal Information*.

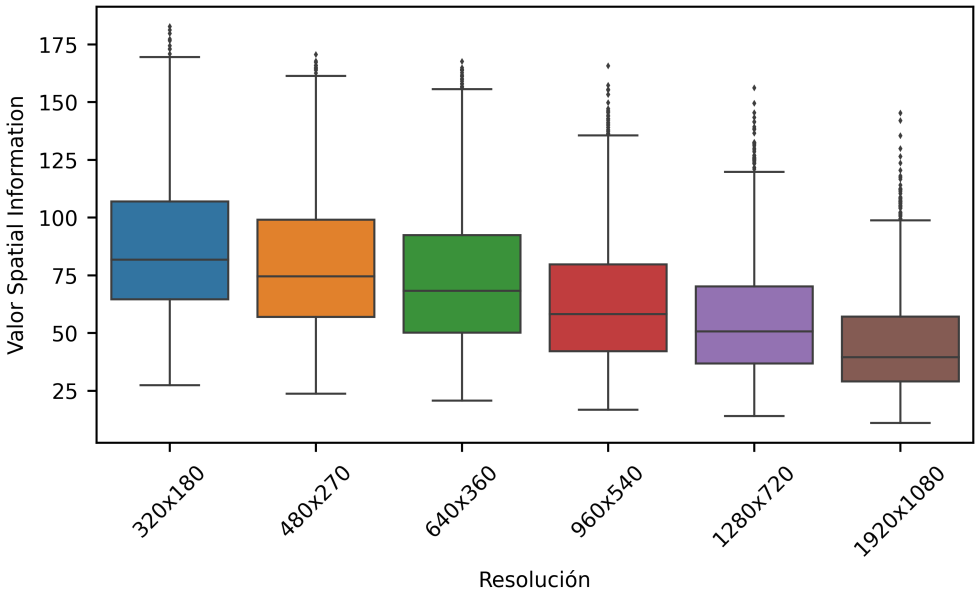


**Figura 4.17:** Métrica *Spatial Information* según la resolución.

El principal inconveniente identificado en el enfoque por cambio de resolución radica en la alteración significativa que se produce en las métricas de vídeo basadas a nivel de la imagen de la solución Video-MOS, cuando se aplican sobre imágenes de distintas resoluciones. En el caso particular de la métrica de vídeo *Spatial Information*, los valores promedio obtenidos sobre el conjunto de datos de prueba evidencian una clara variación en función de la resolución utilizada: 87.09 para la resolución 320x180, 80.19 para 480x270, 74.00 para 640x360, 64.05 para 960x540, 56.08 para 1280x720, y 45.65 para 1920x1080. La Figura 4.17 representa gráficamente la distribución de los valores de la métrica *Spatial Information* para cada una de las resoluciones, mostrando una disminución progresiva del valor conforme aumenta la resolución de la imagen.

El cálculo de la métrica *Spatial Information* sigue la fórmula matemática establecida en la recomendación UIT-T P.910 [108], la cual cuantifica la cantidad de detalle o complejidad espacial en una secuencia de vídeo mediante la desviación estándar de la imagen filtrada con el operador de Sobel [9] [233]. Este operador resalta los bordes y las texturas de la imagen original, de manera que valores elevados de *Spatial Information* indican un mayor nivel de detalle, mientras que valores bajos se asocian a contenidos más simples con menor detalle.

El operador de Sobel es un filtro ampliamente utilizado en el procesamiento de imágenes para la detección de bordes, ya que permite identificar cambios abruptos en la intensidad de los píxeles. Aunque el tamaño estándar del filtro de Sobel es de 3x3, según lo establecido en la recomendación ITU-T P.910 [108], también es posible emplear versiones con dimensiones alternativas, como 1x1, 5x5 o 7x7. Sin embargo, el uso de un tamaño de filtro distinto al valor estándar de 3x3 no mejora la precisión de la métrica *Spatial Information* cuando se aplica sobre imágenes con diferentes resoluciones y se compara con los valores obtenidos con

[9] https://docs.opencv.org/4.11.0/d2/d2c/tutorial_sobel_derivatives.html

la resolución original de 1920x1080 píxeles.

El Anexo D recopila los resultados visuales de la aplicación del operador de Sobel con distintos tamaños de filtro sobre una resolución de 480x270 píxeles. La Figura D.1 presenta la imagen original, mientras que la Figura D.2, Figura D.3, Figura D.4 y Figura D.5 muestran las imágenes resultantes al aplicar el operador de Sobel con los tamaños de filtro de 1x1, 3x3, 5x5 y 7x7, respectivamente. Visualmente, las diferencias son notables, y en términos cuantitativos, los errores calculados en comparación con los valores de la métrica obtenida a resolución original confirman la inestabilidad del cálculo con filtros de diferente tamaño al valor estándar de 3x3. En particular, los errores medios sobre el conjunto de datos de prueba a 480x270 píxeles son de 27.02 % para el tamaño del filtro del operador de Sobel de 1x1, 39.34 % para el tamaño de 3x3, 1047.53 % para el tamaño de 5x5, y 14786.43 % para el tamaño de 7x7.

Estos datos ponen de manifiesto la alta sensibilidad de la métrica *Spatial Information* a los cambios de resolución y a la elección del tamaño del filtro del operador de Sobel. Además, se respalda la recomendación establecida en la norma UIT-T P.910 [108], que sugiere no comparar valores de complejidad espacial entre secuencias de vídeo con diferentes resoluciones. La razón radica en la dependencia del filtro de Sobel y la dependencia de la desviación estándar con respecto a la estructura de los bordes y al aspecto visual de la secuencia original. Por ello, las secuencias de vídeo con mayor resolución tienden a presentar valores menores de complejidad espacial y viceversa, como se ilustra en la Figura 4.17.

No obstante, si el cambio de resolución se aplica exclusivamente sobre las tres métricas de vídeo basadas a nivel de píxel (*Temporal Information*, *Brightness* y *Contrast*), manteniendo el procesamiento del resto de las métricas a resolución original, los valores de error de MOS obtenidos son de 0.04 al emplear resoluciones de 480x270 y 320x180 píxeles. La similitud en los valores de error obtenidos en la estimación de MOS justifica el uso de la resolución de vídeo más baja considerada en esta investigación.

La Tabla 4.15 presenta el tiempo medio empleado por imagen en la extracción de características, con esta aproximación de procesamiento de las métricas basadas a nivel de imagen a resolución original y procesamiento de las métricas basadas a nivel de píxel a baja resolución de 320x180 píxeles. Los datos se han obtenido de la Tabla 4.2, seleccionando el tiempo medio empleado por cada métrica en función de la resolución de la imagen.

El tiempo de procesamiento medio por imagen para la extracción de la información de vídeo, utilizando la aproximación de doble resolución en el enfoque por cambio de resolución, es de 466.42 ms. Esta aproximación implica un ahorro de coste computacional del 14.11 %, al pasar de 543.03 ms a 466.42 ms, en el tiempo medio por imagen para la extracción de las características de vídeo.

El ahorro de coste computacional cercano al 15 % y la obtención de un error de MOS inferior a 0.04 validan el enfoque por cambio de resolución como técnica de óptimización del coste computacional de la herramienta Video-MOS, siempre y cuando se realicen las particularidades descritas anteriormente: procesamiento de las métricas de vídeo *Spatial Information*, *Blurring*, *Ringing*, *Blockloss* y *Blocking* a resolución original de 1920x1080 píxeles; y procesamiento de las métricas *Temporal Information*, *Brightness* y *Contrast* a baja resolución de 320x180 píxeles.

**Tabla 4.15:** Tiempo de procesamiento medio por imagen para las métricas de vídeo con la aproximación de doble resolución.

| **Métrica de evaluación** | **Tiempo medio (ms)** | **Categoría** | **Resolución** |
|---|---|---|---|
| *Spatial Information* | 46.36 | A nivel de imagen | 1920x1080 |
| *Temporal Information* | 0.26 | A nivel de píxel | 320x180 |
| *Blurring* | 20.15 | A nivel de imagen | 1920x1080 |
| *Brightness* | 0.26 | A nivel de píxel | 320x180 |
| *Contrast* | 0.69 | A nivel de píxel | 320x180 |
| *Ringing* | 10.31 | A nivel de imagen | 1920x1080 |
| *Blockloss* | 156.35 | A nivel de imagen | 1920x1080 |
| *Blocking* | 232.04 | A nivel de imagen | 1920x1080 |
| **Total** | **466.42** | - | - |

## 4.2 Enfoques basados en redundancia temporal

En el ámbito del procesamiento de vídeo, la redundancia temporal hace referencia a la repetición o similitud de información entre imágenes consecutivas dentro de una secuencia de vídeo. En la mayoría de los vídeos, los cambios entre una imagen y la siguiente suelen ser mínimos. Este principio es ampliamente utilizado en los estándares de codificación de vídeo, implementando algoritmos diseñados para reducir la cantidad de datos necesarios y almacenando únicamente la información que describe las diferencias entre imágenes consecutivas.

Los enfoques propuestos en esta sección se fundamentan en la redundancia temporal, con el objetivo de reducir el procesamiento de imágenes redundantes dentro de la secuencia de vídeo para optimizar el uso de los recursos computacionales.

### 4.2.1 Enfoque por muestreo temporal uniforme

Una estrategia simple y efectiva para explotar la redundancia temporal en una secuencia de vídeo consiste en seleccionar imágenes específicas dentro de un vídeo, aplicando un muestreo temporal uniforme para procesar únicamente un subconjunto determinado de imágenes. Este enfoque permite reducir significativamente el número de imágenes a procesar para extraer las características sin comprometer notablemente la precisión en la estimación de calidad. El enfoque propuesto considera la posibilidad de procesar desde una de cada dos imágenes hasta únicamente la primera imagen de la secuencia de tres segundos de duración.

Los modos (MOD, nombre heredado de la biblioteca FFmpeg y que hace referencia a una opción del filtro *select* para seleccionar un conjunto determinado de imágenes) definidos en este enfoque por muestreo temporal uniforme son: MOD2, MOD5, MOD10, MOD15, MOD20, MOD25, MOD38 y QO. El modo Q0 representa el caso extremo en el que solo se procesa

**Tabla 4.16:** Error de MOS vs. coste computacional en el enfoque por muestreo temporal uniforme.

| Modo | Imágenes (AVG) | Imágenes (STD) | Coste computacional (%) | Error de MOS (MAE) |
|---|---|---|---|---|
| MOD2 | 37.94 | 1.00 | 50.67 | 0.06 |
| MOD5 | 14.98 | 0.37 | 20.01 | 0.09 |
| MOD10 | 7.99 | 0.20 | 10.67 | 0.12 |
| MOD15 | 5.00 | 0.11 | 6.68 | 0.14 |
| MOD20 | 4.00 | 0.11 | 5.34 | 0.16 |
| MOD25 | 3.00 | 0.08 | 4.01 | 0.18 |
| MOD38 | 2.00 | 0.07 | 2.68 | 0.23 |
| Q0 | 1.00 | 0.07 | 1.34 | 0.39 |

la primera imagen de la secuencia de vídeo. En los modos MODX, la variable X indica la distancia, en número de imágenes, entre dos imágenes consecutivas procesadas dentro de la secuencia. Por ejemplo, el MOD15 implica que se procesa una imagen de cada quince imágenes. Por tanto, en una secuencia de vídeo de tres segundos de duración y a una tasa de refresco de 25 imágenes por segundo, con este modo se procesarían únicamente cinco imágenes (y no las setenta y cinco imágenes procesadas en el modo de funcionamiento normal de la herramienta Video-MOS). Esta reducción en la cantidad de imágenes a procesar supone un ahorro computacional del 93.32 %, representando una optimización significativa en términos de eficiencia computacional.

Para garantizar el correcto funcionamiento de la herramienta Video-MOS utilizando el enfoque por muestreo temporal uniforme, las imágenes no procesadas conservan las mismas características y métricas de vídeo que la última imagen procesada. Con esta decisión se asume que una imagen no procesada es equivalente a la última imagen procesada, y se garantiza la continuidad operativa de la solución sin introducir inconsistencias en los resultados. Cabe destacar que, aunque se realizaron pruebas experimentales utilizando diferentes métodos de interpolación de los valores de las métricas de vídeo entre imágenes procesadas, los resultados no evidenciaron mejoras significativas en la estimación del valor de MOS. Por ello, se ha optado por descartar la interpolación de valores y mantener el mismo valor de las métricas entre imágenes consecutivas no procesadas.

Además, para garantizar que al menos una imagen sea procesada en cada medida de vídeo, independientemente del modo utilizado y de la tasa de refresco del vídeo, se ha tomado la decisión de procesar siempre la primera imagen de la secuencia. Esta decisión asegura la continuidad del análisis y evita casos en los que ninguna imagen sea procesada, lo que podría comprometer la integridad de la estimación de calidad.

La Tabla 4.16 presenta los resultados obtenidos para los diferentes modos propuestos en el enfoque por muestreo temporal uniforme. Se detalla el número de imágenes procesadas por medida de tres segundos (valor medio y desviación estándar), el coste computacional medio (basado en el valor medio de las imágenes procesadas), y el error de MOS (expresado en términos MAE), para cada uno de los modos propuestos. Dependiendo del número de

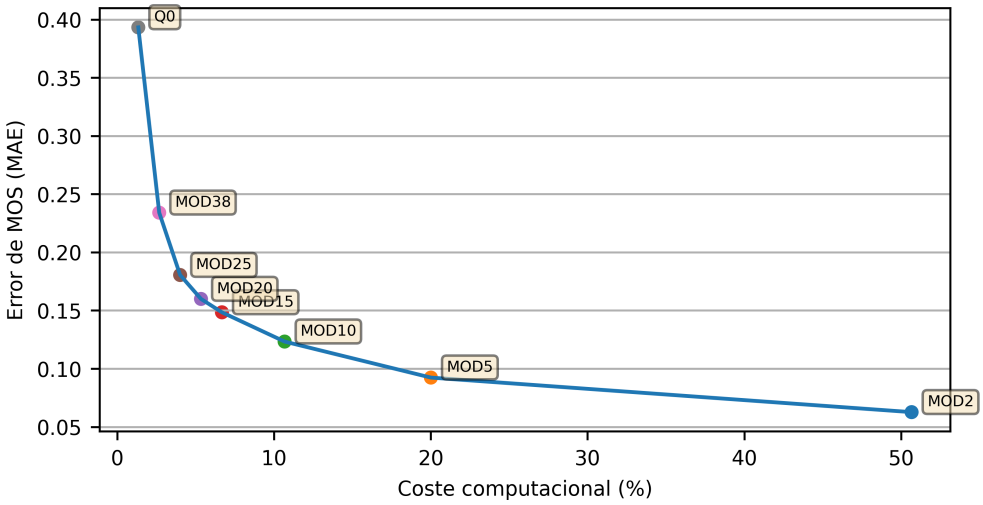


**Figura 4.18:** Error de MOS vs. coste computacional en el enfoque por muestreo temporal uniforme.

imágenes procesadas y utilizadas para la extracción de características, el coste computacional de este enfoque varía entre 1.34 % y 50.67 %, lo que representa un ahorro significativo en comparación con el procesamiento de todas las imágenes de la secuencia.

El error en la estimación del valor de MOS varía entre 0.06 para el modo MOD2 y 0.39 para el modo Q0. La Figura 4.18 ilustra gráficamente la relación entre el coste computacional y el error de MOS para los distintos modos analizados. La curva obtenida muestra una tendencia logarítmica decreciente, que indica que el error en la estimación de calidad disminuye a medida que se incrementa el número de imágenes procesadas dentro de la secuencia de vídeo.

Para garantizar el funcionamiento en tiempo real en el equipo de prueba (ver Sección 3.4), el coste computacional total debe de ser inferior a 7.37 % (ahorro computacional superior al 92.63 %). Un coste computacional de 7.37 % equivale, en promedio, a procesar un máximo de 5.52 imágenes por cada medida de tres segundos de vídeo. Redondeando a cinco imágenes de media por medida, se debería seleccionar un modo MOD15 o superior.

Según la Tabla 4.16, el modo MOD15 ofrece un ahorro del coste computacional del 93.32 %, con un error de MOS de 0.14, cumpliendo tanto el requisito de tiempo real como el requisito de error de MOS inferior a 0.15. En cambio, los modos MOD20, MOD25, MOD38 y Q0 presentan errores de MOS de 0.16, 0.18, 0.23 y 0.39, respectivamente, lo que los descarta por exceder el umbral límite de 0.15.

La Tabla 4.17 también muestra el error cometido en la extracción de características para cada uno de los modos del enfoque por muestreo temporal uniforme. El error del vector de características aumenta a medida que se procesa un menor número de imágenes. Con el modo MOD15, único modo que ha sido seleccionado por cumplir con los requisitos de tiempo real y precisión en la estimación de calidad, el error en el vector de características es de 3.41 %, un valor significativamente menor que los errores obtenidos previamente en los enfoques basados en la redundancia espacial del vídeo.

El enfoque por muestreo temporal uniforme selecciona un número fijo de imágenes a procesar

**Tabla 4.17:** Error en la extracción de características en el enfoque por muestreo temporal uniforme.

| **Métrica de evaluación** | **Error (MAE, en %)** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **MOD2** | **MOD5** | **MOD10** | **MOD15** | **MOD20** | **MOD25** | **MOD38** | **Q0** |
| *Spatial Information* | 0.39 | 0.31 | 0.82 | 1.08 | 1.50 | 1.86 | 2.81 | 5.68 |
| *Temporal Information* | 1.67 | 2.78 | 4.03 | 5.02 | 5.97 | 6.31 | 8.41 | 14.69 |
| *Blurring* | 1.87 | 3.04 | 6.40 | 8.21 | 10.41 | 11.97 | 17.89 | 34.77 |
| *Brightness* | 0.28 | 0.57 | 1.33 | 2.02 | 2.63 | 3.58 | 5.31 | 11.00 |
| *Contrast* | 0.23 | 0.45 | 1.06 | 1.60 | 2.08 | 2.76 | 4.09 | 8.33 |
| *Ringing* | 0.78 | 0.84 | 1.94 | 2.58 | 3.55 | 4.08 | 5.68 | 11.23 |
| *Blockloss* | 0.00 | 0.00 | 1.04 | 1.04 | 2.08 | 2.08 | 2.08 | 2.08 |
| *Blocking* | 3.23 | 2.49 | 4.95 | 5.76 | 7.87 | 8.73 | 11.14 | 17.33 |
| **Vector de caract.** | **1.06** | **1.31** | **2.70** | **3.41** | **4.51** | **5.17** | **7.18** | **13.14** |

en una secuencia de vídeo, sin considerar la complejidad del contenido ni los metadatos de las imágenes. Si bien esto garantiza un funcionamiento en tiempo real al mantener constante el número de imágenes procesadas por medida, presenta una limitación importante: en secuencias simples, se podrían procesar más imágenes de las necesarias, mientras que en secuencias complejas, la cantidad de imágenes procesadas podrían ser insuficientes para una estimación precisa del valor de MOS.

Para mejorar este enfoque, se explorarán otras alternativas que permitan utilizar muestreos temporales más largos al modo MOD15, considerando los modos MOD20, MOD25, MOD38 o Q0. Estas estrategias buscarán reducir inicialmente la cantidad fija de imágenes procesadas y activar el procesamiento adicional cuando sea necesario, en función de un mecanismo adaptativo basado en la complejidad de la secuencia de vídeo.

Para permitir el uso de modos con muestreo temporal más largos sin comprometer la precisión en la estimación de MOS, se proponen en esta investigación dos mecanismos adicionales que activen el procesamiento de imágenes específicas cuando sea necesario:

1. **Uso de la métrica SSIM de vídeo**. Se emplea esta métrica FR para detectar cambios significativos entre imágenes consecutivas. Si la diferencia del valor SSIM entre dos imágenes supera un determinado umbral, se procesará la imagen en cuestión. Este mecanismo permite adaptarse dinámicamente a variaciones en la secuencia de vídeo, evitando la omisión de no procesar imágenes con cambios relevantes.

2. **Uso del tipo de imagen en la codificación de vídeo**. En los estándares de compresión de vídeo, un grupo de imágenes, o estrucutra GOP (*Group Of Pictures*), especifica el orden en que las imágenes son ordenadas. Las imágenes se clasifican en imágenes tipo I (*Intra*), tipo P (*Predicted*) y tipo B (*Bidirectional*). Este mecanismo aprovecha esta información para priorizar el procesamiento de las imágenes de mayor importancia dentro de la secuencia. Por ejemplo, las imágenes tipo I contienen información completa de la imagen y suelen ser más relevantes para el análisis de calidad, por lo que se podría forzar su procesamiento.

Estos dos mecanismos descritos anteriormente permiten ampliar el muestreo temporal del modo MOD15, incorporando una variable de control inteligente que active el procesamiento de imágenes según su importancia dentro de la secuencia de vídeo.

### 4.2.2 Enfoque por muestreo temporal uniforme y métrica SSIM

Este enfoque combina el muestreo temporal uniforme con la métrica SSIM, permitiendo seleccionar un modo de muestreo temporal más largo que el modo MOD15, con el objetivo de reducir la cantidad de imágenes procesadas de manera fija. Para garantizar el funcionamiento correcto de la solución Video-MOS con este enfoque, se mantiene la obligación de procesar siempre la primera imagen de cada medida de vídeo.

El mecanismo basado en la métrica SSIM compara cada imagen dentro de una secuencia de vídeo con la última imagen procesada. Si la diferencia entre ambas supera un umbral predefinido, se activa la condición de procesamiento, lo que permite extraer las características de dicha imagen específica. De este modo, solo se procesan imágenes adicionales cuando se detectan cambios significativos que puedan modificar las características utilizadas en la estimación de calidad, optimizando así el uso de los recursos computacionales.

Una particularidad del enfoque por muestreo temporal uniforme y métrica SSIM radica en el cálculo de la métrica de vídeo *Temporal Information*. Es imprescindible calcular esta métrica en la imagen inmediatamente posterior a cualquier imagen procesada debido a la activación por SSIM. De no realizarse el cálculo, la diferencia detectada se mantendría en la métrica *Temporal Information* en las imágenes siguientes hasta que una nueva imagen fuera procesada. Este ajuste garantiza que la métrica *Temporal Information* refleje correctamente los cambios temporales en la secuencia de vídeo, evitando inconsistencias en los valores obtenidos y asegurando una evaluación precisa de la variabilidad temporal en la calidad percibida.

El hecho de que el valor de SSIM esté normalizado en un rango de 0 a 1 permite una interpretación clara y directa del grado de similitud entre las imágenes comparadas. Valores próximos a 1 indican una alta similitud, lo que sugiere que la imagen en cuestión mantiene la estructura y las características visuales de la última imagen procesada. En cambio, valores más bajos de SSIM reflejan diferencias notables, ocasionados por variaciones en la iluminación, movimientos u otros cambios en la estructura de la imagen. Esta propiedad es fundamental dentro de este enfoque propuesto, ya que es necesario definir un umbral que regule la decisión de procesar una imagen o descartarla en función de su similitud con la última imagen procesada, sin comprometer la precisión del análisis.

El umbral de la métrica SSIM en la detección de cambios entre imágenes consecutivas juega un papel crucial en la determinación del número de imágenes a procesar, incidiendo directamente en el coste computacional del enfoque propuesto. Un umbral bajo, cercano a 0, permite reducir el número de imágenes procesadas, dado que solo se identificarán cambios significativos entre ellas. Sin embargo, un umbral elevado cercano a 1, incrementaría sustancialmente el número de imágenes procesadas, lo que derivaría en un coste computacional excesivo y podría comprometer la eficiencia del enfoque. La correcta selección del umbral SSIM es, por tanto, un aspecto clave en la optimización del balance entre precisión en la estimación de calidad y eficiencia computacional.

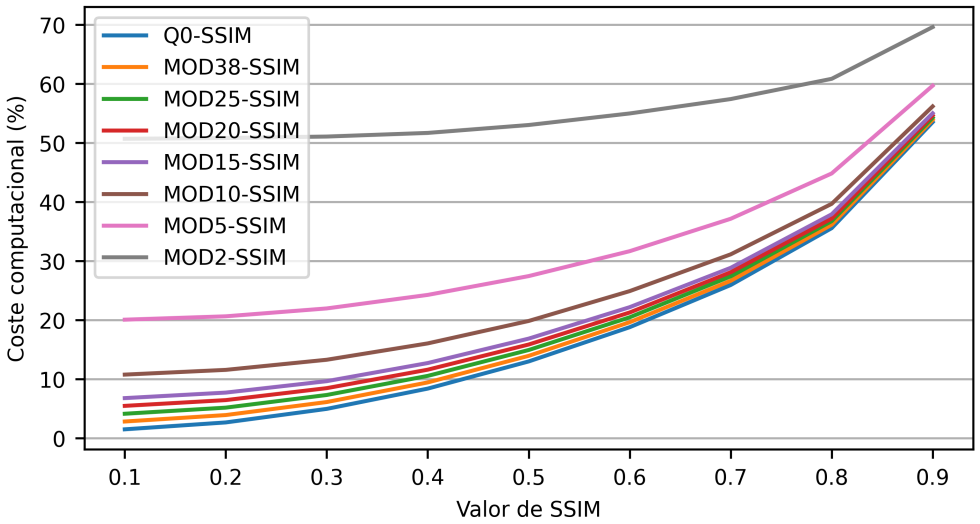


**Figura 4.19:** Coste computacional vs. valor de SSIM en el enfoque por muestreo temporal uniforme y métrica SSIM.

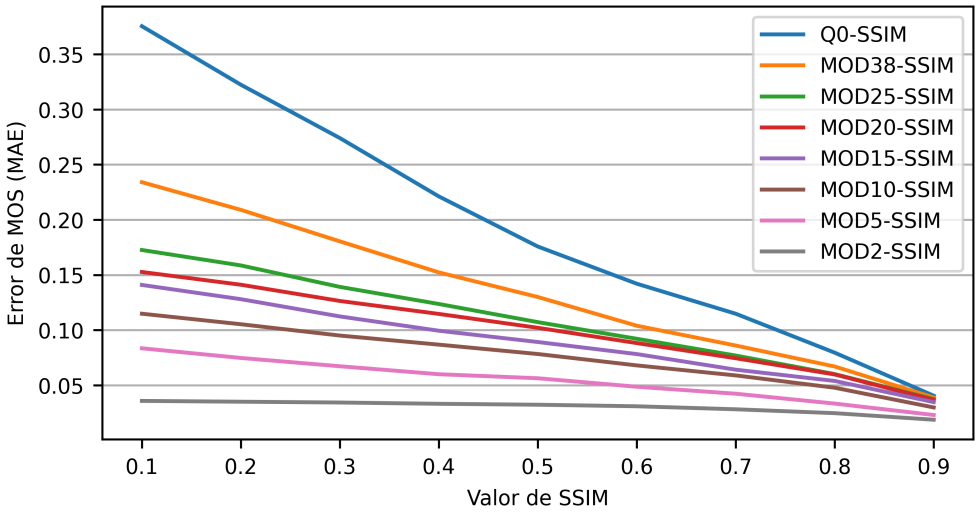


**Figura 4.20:** Error de MOS vs. valor de SSIM en el enfoque por muestreo temporal uniforme y métrica SSIM.

Los resultados obtenidos al aplicar nueve valores diferentes de umbral SSIM, en un rango de 0.1 a 0.9, para los ocho modos considerados en el enfoque por muestreo temporal uniforme, evidencian la relación existen entre el umbral SSIM, el coste computacional y la precisión en la estimación del valor de MOS. La representación gráfica de los resultados permite visualizar cómo el coste computacional varía en función del umbral SSIM (Figura 4.19) y cómo esta variación influye notablemente en el error de MOS obtenido (Figura 4.20).

El análisis de los resultados ha permitido establecer que el valor máximo del umbral SSIM que garantiza funcionamiento en tiempo real para los modos MOD20, MOD25, MOD38 y

**Tabla 4.18:** Error de MOS vs. coste computacional en el enfoque por muestreo temporal uniforme y métrica SSIM.

| Modo | Imágenes (AVG) | Imágenes (STD) | Coste computacional (%) | Error de MOS (MAE) |
|---|---|---|---|---|
| MOD2-SSIM03 | 38.24 | 1.74 | 51.07 | 0.03 |
| MOD5-SSIM03 | 16.46 | 4.56 | 21.98 | 0.07 |
| MOD10-SSIM03 | 9.95 | 5.28 | 13.29 | 0.10 |
| MOD15-SSIM03 | 7.24 | 5.82 | 9.67 | 0.11 |
| MOD20-SSIM03 | 6.35 | 5.90 | 8.48 | 0.13 |
| MOD25-SSIM03 | 5.49 | 6.14 | 7.33 | 0.14 |
| MOD38-SSIM03 | 4.59 | 6.25 | 6.13 | 0.18 |
| Q0-SSIM03 | 3.72 | 6.43 | 4.97 | 0.27 |

Q0 es 0.3. La elección de este umbral responde a la necesidad de buscar un equilibrio entre la reducción de coste computacional y la precisión en la estimación de MOS. En términos generales, un umbral SSIM de 0.3 permite mantener el rendimiento en tiempo real de la herramienta Video-MOS en el equipo de prueba, asegurando además un error de MOS inferior a 0.15, lo que satisface los requisitos definidos para la optimización de la solución.

Los resultados obtenidos con este umbral de SSIM de 0.3 se presentan en la Tabla 4.18, donde se detalla el número de imágenes procesadas, el coste computacional y el error de MOS para cada uno de los modos propuestos en el enfoque. Los resultados correspondientes al resto de valores umbral de SSIM se presentan en el Anexo E.

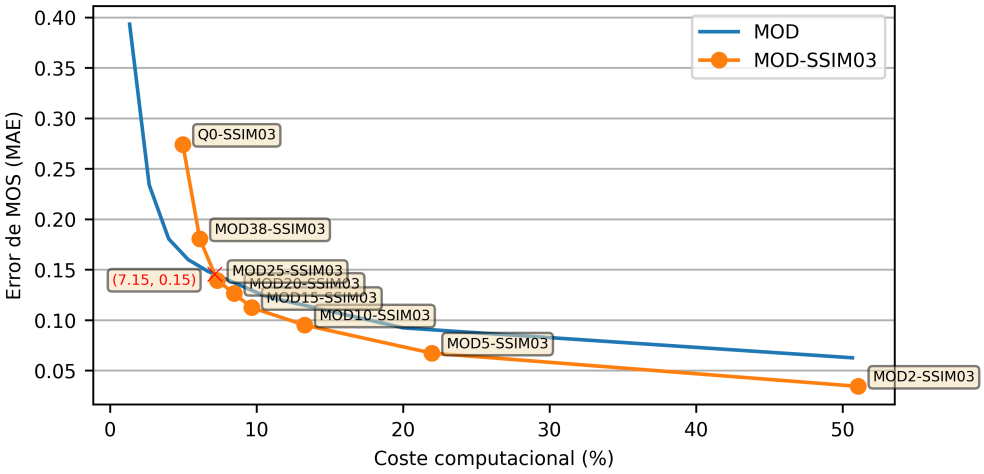


**Figura 4.21:** Error de MOS vs. coste computacional en el enfoque por muestreo temporal uniforme y métrica SSIM.

El gráfico que representa la relación entre el coste computacional y el error de MOS para este enfoque propuesto se muestra en la Figura 4.21. En esta representación también se incluye la

**Tabla 4.19:** Error en la extracción de características en el enfoque por muestreo temporal uniforme y métrica SSIM.

| Métrica de evaluación | Error (MAE, en %) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | MOD2-SSIM03 | MOD5-SSIM03 | MOD10-SSIM03 | MOD15-SSIM03 | MOD20-SSIM03 | MOD25-SSIM03 | MOD38-SSIM03 | Q0-SSIM03 |
| *Spatial Information* | 0.37 | 0.21 | 0.59 | 0.67 | 0.95 | 1.08 | 1.56 | 2.59 |
| *Temporal Information* | 0.00 | 1.05 | 2.05 | 2.58 | 3.53 | 3.96 | 5.61 | 8.25 |
| *Blurring* | 1.79 | 2.66 | 5.09 | 5.61 | 6.89 | 7.43 | 9.44 | 13.01 |
| *Brightness* | 0.24 | 0.36 | 0.81 | 1.14 | 1.48 | 1.91 | 2.73 | 4.74 |
| *Contrast* | 0.20 | 0.31 | 0.68 | 0.99 | 1.25 | 1.60 | 2.37 | 4.03 |
| *Ringing* | 0.74 | 0.72 | 1.56 | 1.95 | 2.57 | 2.78 | 3.88 | 6.02 |
| *Blockloss* | 0.50 | 0.89 | 1.30 | 1.49 | 1.62 | 1.95 | 2.15 | 2.33 |
| *Blocking* | 3.15 | 2.48 | 4.90 | 5.90 | 7.97 | 9.00 | 11.87 | 18.27 |
| **Vector de caract.** | **0.87** | **1.09** | **2.12** | **2.54** | **3.28** | **3.72** | **4.95** | **7.41** |

curva correspondiente al enfoque por muestreo temporal uniforme (ver Figura 4.18), lo que permite establecer un punto de referencia para la comparación. Se observa que la incorporación de la métrica SSIM aporta una mejora en los resultados cuando el coste computacional supera el 7.15 %, que corresponder con el punto de intersección entre ambas curvas.

El comportamiento de la gráfica refleja dos tramos claramente diferenciados. En el primer tramo, cuando el coste computacional del enfoque es inferior a 7.15 %, la incorporación de la métrica SSIM tiende a disminuir la precisión en la estimación del valor de MOS en comparación con el enfoque por muestreo temporal uniforme estándar. Sin embargo, a partir del umbral del 7.15 %, el enfoque con SSIM mejora la precisión de la estimación, reduciendo el error del valor de MOS para un mismo número de imágenes procesadas.

En cuanto a los errores obtenidos en las métricas de vídeo de la solución Video-MOS y en el vector de características, los resultados se presentan en la Tabla 4.19. Se aprecia que los errores en el vector de características son relativamente bajos, oscilando entre 0.87 % en el modo MOD2-SSIM03 y 7.41 % en el modo Q0-SSIM03. Como era de esperar, la reducción del número de imágenes procesadas conlleva un incremento en el error, lo que confirma la influencia directa de la cantidad de datos analizados en la precisión de la estimación de calidad.

El modo seleccionado para la implementación del enfoque por muestreo temporal uniforme y métrica SSIM es el modo MOD25-SSIM03, que presenta un coste computacional medio del 7.33 % y un ahorro del 92.67 % en comparación con el modo de funcionamiento normal de la herramienta. Además, este modo ofrece un error en la estimación del valor de MOS de 0.14, manteniéndose por debajo del umbral de 0.15, establecido como indicativo de precisión en la evaluación de calidad de vídeo.

Si bien el coste computacional medio permitiría el funcionamiento en tiempo real de la solución Video-MOS, procesando en promedio 5.49 imágenes por cada medida de tres segundos de vídeo, la elevada variabilidad en el número de imágenes procesadas representa una limitación significativa en cuanto a garantizar el procesamiento en tiempo real. En particular, la desviación

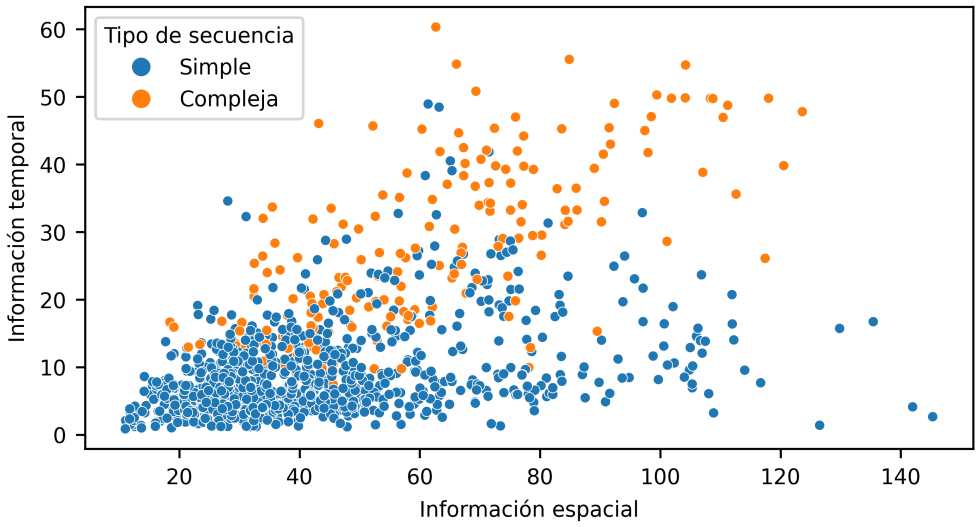


**Figura 4.22:** Diagrama SI-TI de las secuencias de prueba. Identificación de secuencias complejas.

estándar para el modo MOD25-SSIM03 alcanza un valor de 6.14 imágenes, lo que refleja una dispersión considerable en la cantidad de imágenes procesadas en función de la complejidad y tipo de vídeo.

Esta limitación se evidencia en el 17.36 % de las secuencias de vídeo del conjunto de prueba (195 secuencias), donde el modo MOD25-SSIM03 no cumple con el requisito de procesamiento en tiempo real al superar el umbral de coste computacional permitido. La Figura 4.22 muestra la distribución de estas 195 secuencias en el diagrama SI-TI del conjunto de prueba (ver Figura 3.4), resaltando la elevada complejidad de dichas secuencias, caracterizadas por cambios significativos entre imágenes consecutivas y altos valores de complejidad temporal.

Con el objetivo de garantizar el funcionamiento en tiempo real, independientemente del tipo de secuencia de vídeo, se propone una versión limitada del modo MOD25-SSIM03, denominada MOD25-SSIM03-LIM5. Esta versión impone un umbral máximo de cinco imágenes procesadas por medida para controlar el coste computacional. De este modo, se asegura que este enfoque mantenga un rendimiento computacional estable, sin verse afectado por la variabilidad o la complejidad de las secuencias de vídeo.

Los resultados obtenidos para los diferentes modos de funcionamiento (MOD15, MOD25-SSIM03 y MOD25-SSIM-LIM5) se presentan en la Tabla 4.20 y la Tabla 4.21. La Tabla 4.20 recoge los datos correspondientes al conjunto total de secuencias de prueba, mientras que la Tabla 4.21 proporciona la información relacionada con el subconjunto de secuencias más complejas, compuesto por las 195 secuencias que no permitían el funcionamiento en tiempo real del modo MOD25-SSIM03.

Para el conjunto total de secuencias de prueba, los resultados obtenidos muestran una ligera mejora con el modo MOD25-SSIM03-LIM5 en comparación con el modo MOD15. Ambos modos permiten el funcionamiento en tiempo real de la herramienta, con errores de MOS de 0.16 y 0.14, respectivamente, y costes computacionales de 4.94 % y 6.68 %.

**Tabla 4.20:** Error de MOS vs. coste computacional en el enfoque por muestreo temporal uniforme y métrica SSIM. 100 % de las secuencias de prueba.

| Modo | Imágenes (AVG) | Imágenes (STD) | Coste computacional ( %) | Error de MOS (MAE) |
|---|---|---|---|---|
| MOD15 | 5.00 | 0.11 | 6.68 | 0.14 |
| MOD25-SSIM03 | 5.49 | 6.14 | 7.33 | 0.14 |
| MOD25-SSIM03-LIM5 | 3.70 | 0.83 | 4.94 | 0.16 |

**Tabla 4.21:** Error de MOS vs. coste computacional en el enfoque por muestreo temporal uniforme y métrica SSIM. 17.36 % de las secuencias de prueba.

| Modo | Imágenes (AVG) | Imágenes (STD) | Coste computacional ( %) | Error de MOS (MAE) |
|---|---|---|---|---|
| MOD15 | 5.01 | 0.10 | 6.68 | 0.18 |
| MOD25-SSIM03 | 15.30 | 9.94 | 20.40 | 0.11 |
| MOD25-SSIM03-LIM5 | 5.00 | 0.00 | 6.67 | 0.22 |

Aunque los errores de MOS son prácticamente similares entre ambos modos, el coste computacional es ligeramente inferior para el modo MOD25-SSIM03-LIM5, lo que sugiere una mayor eficiencia sin comprometer excesivamente la precisión.

No obstante, al analizar el conjunto de secuencias más complejas (17.36 %), los resultados muestran diferencias importantes. El modo MOD25-SSIM03-LIM5, con un coste computacional de 6.67 % obtiene un error de MOS de 0.22, mientras que el modo MOD15, con un coste computacional equivalente, presenta un error de MOS algo menor de 0.18. Este comportamiento indica que, en caso de secuencias complejas, un enfoque que distribuya de manera uniforme las imágenes a procesar resulta ser más efectivo que el uso de la métrica SSIM. En este contexto, si se acepta renunciar al requisito de funcionamiento en tiempo real, el modo MOD25-SSIM03, con un coste computacional medio de 20.40 %, proporciona la estimación más precisa del valor de MOS, con un valor de 0.11. Este resultado evidencia que un incremento en el número de imágenes procesadas contribuye a una mejora significativa en la precisión de la estimación del valor de MOS, especialmente en secuencias de vídeo con mayor complejidad.

Si se busca garantizar el funcionamiento en tiempo real, el enfoque por muestreo temporal uniforme y métrica SSIM no presenta ventajas significativas en comparación con el enfoque por muestreo temporal uniforme estándar. En particular, el modo MOD25-SSIM03-LIM5 no proporciona una reducción sustancial del coste computacional ni una mejora notable en la precisión de la estimación del valor de MOS en comparación con el modo MOD15.

Además, la implementación del mecanismo basado en SSIM introduce un coste computacional adicional debido al cálculo de esta métrica en todas las imágenes de la medida de vídeo. Las pruebas realizadas con más de 84000 imágenes de prueba han constatado el elevado coste computacional asociado, el cual aumenta proporcionalmente con la resolución de la imagen. En particular, el tiempo medio por imagen para el cálculo de SSIM es de: 290.21 ms para

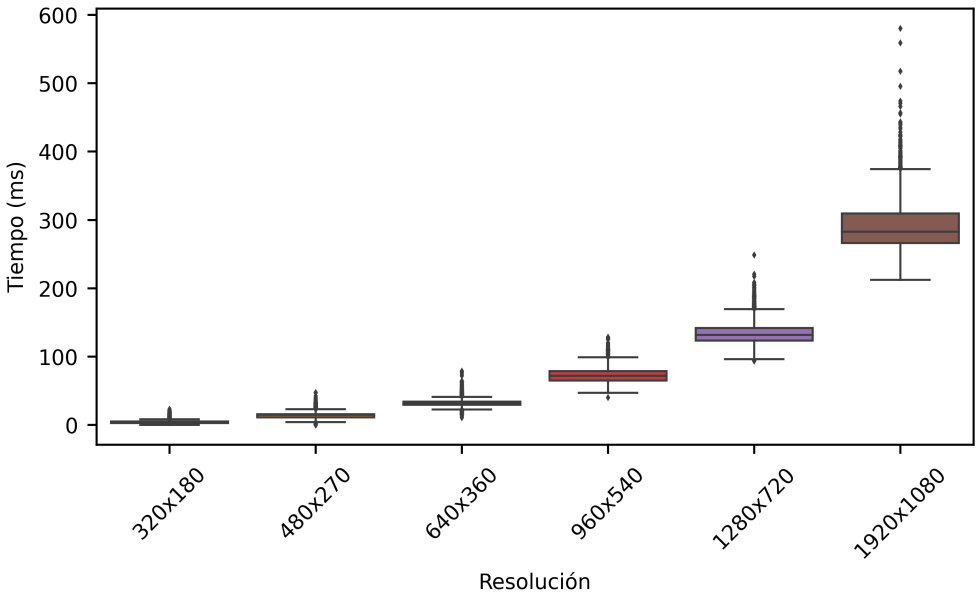


**Figura 4.23:** Tiempo medio de cálculo de SSIM por imagen según la resolución.

1920x1080, 135.14 ms para 1280x720, 73.31 ms para 960x540, 32.76 ms para 640x360, 14.49 ms para 480x270 y 4.58 ms para 320x180.

La Figura 4.23 muestra la relación entre el tiempo empleado por imagen en el cálculo de SSIM y la resolución. Si se analiza una medida de tres segundos de duración con una tasa de refresco de 25 imágenes por segundo, el tiempo total requerido para calcular la métrica SSIM en todas las imágenes de la medida, a una resolución de 320x180 píxeles, asciende aproximadamente a 343.35 ms. Este valor (11.45 % del tiempo de medida) representa un impacto significativo en el rendimiento de la solución basada en este enfoque, evidenciando el coste adicional introducido por el cálculo de la métrica SSIM incluso a resoluciones bajas.

Los inconvenientes derivados del uso de la métrica SSIM para garantizar el funcionamiento en tiempo real de la herramienta Video-MOS, sumados al coste computacional adicional asociado a su cálculo, ponen de manifiesto que el enfoque por muestreo temporal uniforme y métrica SSIM no supone una alternativa viable para la optimización del rendimiento computacional de la solución Video-MOS.

### 4.2.3 Enfoque por muestreo temporal uniforme y tipo de imagen

El enfoque por muestreo temporal uniforme y tipo de imagen cambia el uso de la métrica SSIM por metadatos de las imágenes generados en el proceso de codificación del vídeo. Los codificadores de vídeo, como el estándar H.264/AVC [71], empleado mayoritariamente en la actualidad en España para la transmisión de señal TDT, estructuran la codificación de vídeo en tres tipos de imágenes diferentes: tipo I (Intra), tipo P (Predictivo) y tipo B (Bidireccional). Las imágenes tipo I se codifican exclusivamente mediante la predicción *intra-frame* y sirven como referencia para la predicción de imágenes tipo P y tipo B. Por otro lado, las imágenes tipo P y tipo B incorporan predicción *inter-frame*, donde las primeras utilizan únicamente

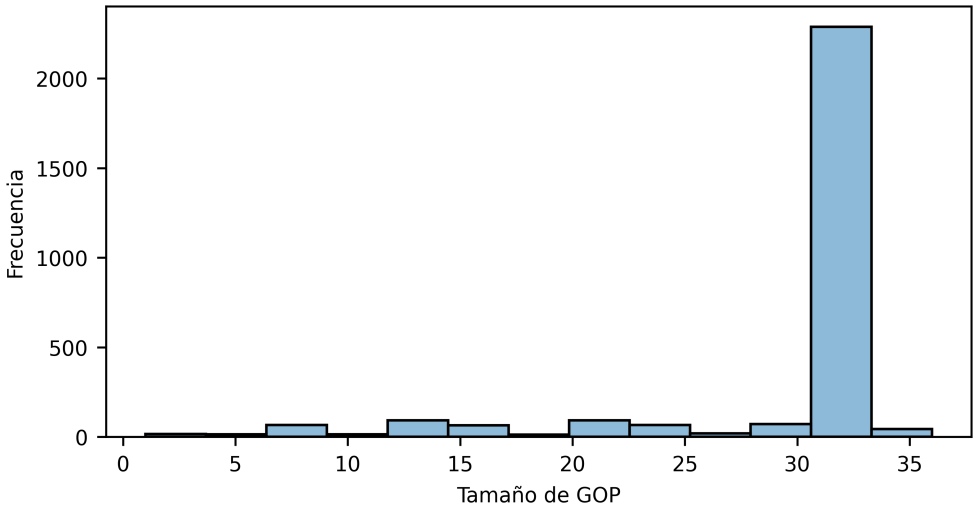


**Figura 4.24:** Tamaños de los GOP en las secuencias de prueba.

imágenes pasadas como referencia, mientras que las segundas pueden emplean tanto imágenes previas como imágenes futuras para mejorar la eficiencia de la predicción.

El estándar H.264/AVC permite la utilización de un tamaño fijo de GOP (Group of Pictures) para la codificación de las secuencias de vídeo. Sin embargo, lo más habitual son las estructuras adaptativas de GOP. Una configuración adaptativa (tamaño variable) de GOP resulta más eficaz en la gestión de cambios de escena y en variaciones significativas entre imágenes consecutivas, optimizando así el proceso de predicción y compresión de la señal. En situaciones donde se detectan transiciones abruptas en el contenido visual, los codificadores que implementan estructuras de GOP adaptativas tienden a insertar imágenes tipo I [234], [235]. La selección del tipo de imagen y la configuración del tamaño del GOP son factores determinantes en el rendimiento del proceso de codificación, tanto en términos de compresión como de calidad del vídeo resultante.

Dado que las imágenes tipo I suelen utilizarse en los cambios de escena, pueden asociarse a instantes caracterizados por una baja redundancia temporal. De la misma manera, imágenes tipo P y tipo B están vinculadas a una mayor redundancia temporal que las imágenes tipo I, al utilizar predicción *inter-frame*. En este sentido, es posible obtener información similar a la proporcionada por la métrica FR SSIM simplemente analizando la estructura de GOP, sin necesidad de calcular explícitamente la similitud entre imágenes consecutivas.

Este nuevo enfoque busca procesar, además de las imágenes fijas seleccionadas por el muestreo temporal uniforme, todas las imágenes tipo I de la secuencia de vídeo. La obtención del tipo de imagen a través de la lectura de metadatos es un proceso prácticamente instantáneo y no conlleva ningún coste computacional adicional. No obstante, el principal desafío de esta estrategia radica en la posible aparición de estructuras de GOP de tamaño reducido, que derive en una cantidad elevada de imágenes tipo I en medidas de tres segundos de vídeo. Esta limitación (similar a la del enfoque basado en métrica SSIM) incrementaría considerablemente el coste computacional más allá del umbral aceptable para el procesamiento en tiempo real,

**Tabla 4.22:** Error de MOS vs. coste computacional en el enfoque por muestreo temporal uniforme y tipo de imagen.

| Modo | Imágenes (AVG) | Imágenes (STD) | Coste computacional (%) | Error de MOS (MAE) |
|---|---|---|---|---|
| MOD2-I | 39.21 | 1.64 | 52.36 | 0.03 |
| MOD5-I | 17.02 | 0.84 | 22.73 | 0.07 |
| MOD10-I | 10.28 | 0.74 | 13.73 | 0.09 |
| MOD15-I | 7.38 | 0.73 | 9.85 | 0.11 |
| MOD20-I | 6.41 | 0.67 | 8.56 | 0.12 |
| MOD25-I | 5.44 | 0.67 | 7.27 | 0.13 |
| MOD38-I | 4.49 | 0.60 | 6.00 | 0.13 |
| Q0-I | 3.52 | 0.60 | 4.70 | 0.14 |

comprometiendo la viabilidad del enfoque.

El análisis del tipo de imagen en el conjunto de secuencias de prueba revela una distribución promedio de 2.56 imágenes tipo I, 15.96 imágenes tipo P y 56.36 imágenes tipo B en medidas de tres segundos de vídeo. En este contexto, la media de 2.56 imágenes tipo I por medida si que permite considerar la posibilidad de procesar siempre este tipo de imágenes, sin comprometer el funcionamiento en tiempo real de la solución Video-MOS. La Figura 4.24 muestra la distribución de los tamaños de GOP en las secuencias analizadas dentro del conjunto de prueba. En total, en las 1123 secuencias de vídeo se evaluaron un total de 2856 GOPs, de los cuales el 66.67 % presenta una estructura $IBBBP$, con parámetros $M = 4$ y $N = 32$. El parámetro $M$ representa la distancia entre una imagen tipo I y la imagen tipo P más próxima, o bien la separación entre dos imágenes tipo P consecutivas. Por su parte, $N$ define el tamaño del GOP, es decir, la distancia entre dos imágenes tipo I consecutivas.

El enfoque por muestreo temporal uniforme y tipo de imagen continúa procesando siempre la primera imagen de la medida de vídeo. De manera análoga al enfoque por muestreo temporal uniforme y métrica SSIM, y dado que la presencia de una imagen tipo I puede ir asociado a un cambio de escena, la métrica de vídeo *Temporal Information* se calcula sobre la imagen inmediatamente posterior a una imagen tipo I.

La Tabla 4.22 presenta los resultados obtenidos con este enfoque. En términos de coste computacional medio, los valores oscilan entre un 4.70 % en el modo Q0-I y un 52.36 % en el modo MOD2-I. El aumento del coste computacional y por tanto el número de imágenes procesadas se manifiesta en una disminución notable del error de MOS, de 0.14 en el modo Q0-I a 0.03 en el modo MOD2-I.

La Figura 4.25 presenta la relación entre el coste computacional y el error de valor de MOS del enfoque por muestreo temporal uniforme y tipo de imagen. Con el objetivo de proporcionar un punto de referencia, el gráfico también incluye la curva correspondiente al enfoque por muestreo temporal uniforme estándar.

La incorporación de imágenes tipo I en la extracción de características de vídeo mejora

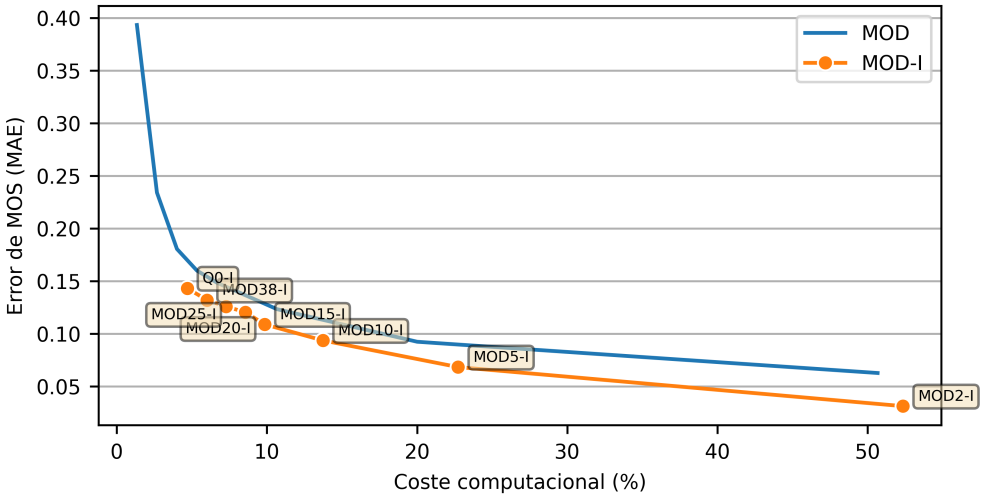


**Figura 4.25:** Error de MOS vs. coste computacional en el enfoque por muestreo temporal uniforme y tipo de imagen.

significativamente los resultados obtenidos en comparación con el enfoque por muestreo temporal uniforme. Esta mejoría se observa independientemente del coste computacional y, por tanto, del número de imágenes procesadas. Para un mismo número de imágenes procesadas, el uso de imágenes tipo I en este enfoque permite reducir el error asociado a la estimación de MOS.

La Tabla 4.23 presenta los errores obtenidos en las distintas métricas de vídeo y en el vector de características. En particular, el error en el vector de características varía entre un 0.81 % en el modo MOD2-I hasta un 8.13 % en el modo Q0-I.

**Tabla 4.23:** Error en la extracción de características en el enfoque por muestreo temporal uniforme y tipo de imagen.

| **Métrica de evaluación** | **Error (MAE, en %)** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **MOD2-I** | **MOD5-I** | **MOD10-I** | **MOD15-I** | **MOD20-I** | **MOD25-I** | **MOD38-I** | **Q0-I** |
| *Spatial Information* | 0.35 | 0.23 | 0.55 | 0.66 | 0.83 | 0.99 | 1.24 | 1.52 |
| *Temporal Information* | 0.00 | 1.13 | 1.89 | 2.44 | 2.91 | 3.34 | 4.08 | 4.70 |
| *Blurring* | 1.77 | 2.67 | 4.87 | 6.83 | 8.01 | 8.79 | 11.38 | 13.04 |
| *Brightness* | 0.24 | 0.37 | 0.70 | 0.98 | 1.14 | 1.40 | 1.71 | 2.04 |
| *Contrast* | 0.20 | 0.32 | 0.61 | 0.87 | 1.01 | 1.26 | 1.53 | 1.84 |
| *Ringing* | 0.72 | 0.70 | 1.40 | 1.86 | 2.17 | 2.51 | 2.96 | 3.54 |
| *Blockloss* | 0.48 | 0.95 | 1.35 | 1.45 | 1.77 | 1.78 | 1.87 | 1.95 |
| *Blocking* | 2.72 | 3.72 | 7.68 | 12.02 | 15.46 | 19.59 | 27.78 | 36.40 |
| **Vector de caract.** | **0.81** | **1.26** | **2.38** | **3.39** | **4.16** | **4.96** | **6.57** | **8.13** |

De acuerdo con los datos presentados en la Tabla 4.22 y la Tabla 4.23, el modo Q0-I proporciona resultados óptimos, con un coste computacional medio de 4.70 % (equivalente a un ahorro computacional del 95.30 %) y un error de MOS inferior a 0.15. En este modo, se procesan en

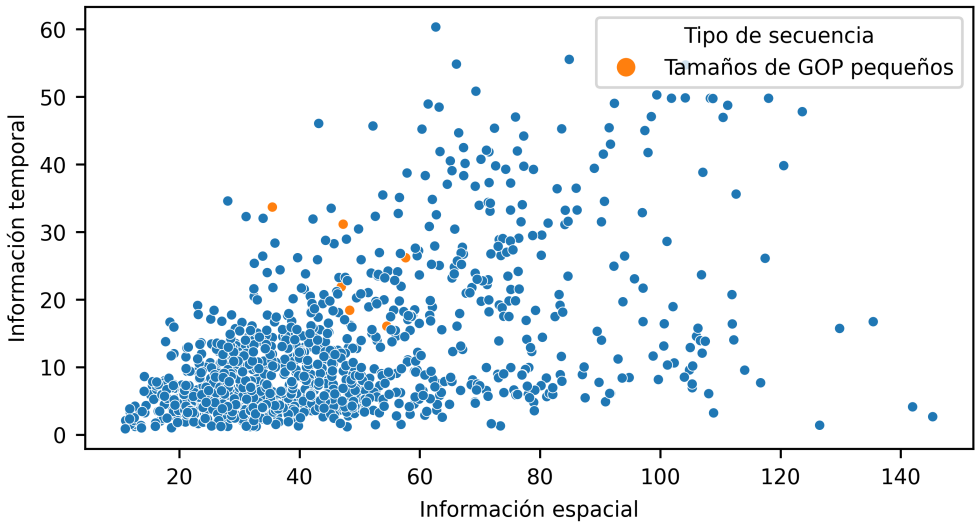


**Figura 4.26:** Diagrama SI-TI de las secuencias de prueba. Identificación de secuencias con tamaños de GOP pequeños.

promedio 3.52 imágenes por medida de tres segundos de vídeo, con una desviación estándar de 0.60.

Un análisis detallado del desempeño del modo Q0-I revela que únicamente el 0.53 % de las secuencias del conjunto de prueba (6 de 1123) no cumpliría con los requisitos de procesamiento en tiempo real al exceder el número máximo de imágenes permitidas. La Figura 4.26 muestra estas seis secuencias identificadas en el diagrama SI-TI, las cuales forman parte también del 17.36 % del conjunto total de medidas clasificadas como secuencias más complejas, según la categorización establecida en la Figura 4.22.

La limitación en el procesamiento en tiempo real se debe a la presencia de tamaños de GOP reducidos en estas seis secuencias, lo que conlleva un mayor número de imágenes tipo I y supera el umbral de cinco imágenes por medida. Aunque la incidencia de estos casos es baja, con el fin de garantizar el funcionamiento en tiempo real independientemente del tipo de secuencia y de los tamaños de GOP, se propone una versión optimizada del modo Q0-I, denominada Q0-I-LIM5. Esta versión establece un límite máximo de cinco imágenes procesadas por medida, asegurando así la viabilidad computacional del modo propuesto.

La Tabla 4.24 y la Tabla 4.25 presentan los resultados obtenidos con el modo Q0-I-LIM5, tanto para el total de secuencias de prueba como para el 17.36 % de las secuencias clasificadas en esta investigación como secuencias de mayor complejidad. Dado que el 0.53 % de las secuencias de vídeo representa un conjunto de datos demasiado reducido, se ha optado por seleccionar el conjunto correspondiente al 17.36 % de las secuencias.

Los resultados obtenidos al comparar el modo Q0-I-LIM5 con el modo MOD15 demuestran que el primero es un modo más eficiente, ya que aprovecha de manera más efectiva la explotación de la redundancia temporal del vídeo. Además, el modo Q0-I-LIM5 no solo ofrece un mayor ahorro computacional, sino también una mejor precisión en la estimación del valor de MOS en ambos conjuntos de datos analizados.

**Tabla 4.24:** Error de MOS vs. coste computacional en el enfoque por muestreo temporal uniforme y tipo de imagen. 100 % de las secuencias de prueba.

| Modo | Imágenes (AVG) | Imágenes (STD) | Coste computacional ( %) | Error de MOS (MAE) |
|---|---|---|---|---|
| MOD15 | 5.00 | 0.11 | 6.68 | 0.14 |
| Q0-I | 3.52 | 0.60 | 4.70 | 0.14 |
| Q0-I-LIM5 | 3.51 | 0.55 | 4.68 | 0.14 |

**Tabla 4.25:** Error de MOS vs. coste computacional en el enfoque por muestreo temporal uniforme y tipo de imagen. 17.36 % de las secuencias de prueba.

| Modo | Imágenes (AVG) | Imágenes (STD) | Coste computacional ( %) | Error de MOS (MAE) |
|---|---|---|---|---|
| MOD15 | 5.01 | 0.10 | 6.68 | 0.18 |
| Q0-I | 3.57 | 0.81 | 4.76 | 0.17 |
| Q0-I-LIM5 | 3.51 | 0.60 | 4.68 | 0.17 |

Para el 17.36 % de las secuencias de prueba más complejas, el modo MOD15 registra un coste computacional medio de 6.68 % y un error de MOS de 0.18, mientras que el modo Q0-I-LIM5 reduce estos valores a 4.68 % y a 0.17, respectivamente. De manera similar, para el conjunto total de prueba, el modo MOD15 presenta un coste computacional medio de 6.68 % y un error de MOS de 0.14, mientras que el modo Q0-I-LIM5 alcanza un coste medio inferior de 4.68 %, y un error de MOS, de 0.14.

Por todo ello, el enfoque por muestreo temporal uniforme y tipo de imagen, en su versión limitada para garantizar el procesamiento en tiempo real en cualquier situación, se consolida como la estrategia más eficaz para aprovechar la redundancia temporal del vídeo. Asimismo, se presenta como una solución óptima para la optimización del coste computacional de la herramienta Video-MOS, asegurando el equilibrio buscado entre rendimiento y precisión.

## 4.3 Enfoques basados en redundancia espacial y temporal

En la sección final de este capítulo se sintetizan las lecciones aprendidas del análisis de las diferentes estrategias utilizadas en la investigación, para aprovechar la redundancia espacial y temporal en secuencias de vídeo. Además, se presenta el enfoque final propuesto para optimizar la herramienta Video-MOS, con un modo que integra los dos tipos de redundancia.

En el análisis de la redundancia espacial en secuencias de vídeo, se propone optimizar el procesamiento de las métricas de vídeo de la herramienta Video-MOS mediante el uso de resoluciones de imagen de menor tamaño, especialmente para aquellas métricas que trabajan

a nivel de píxel. Se establece que las métricas de vídeo: *Temporal Information*, *Brightness* y *Contrast*, se procesen a una resolución de 320x180 píxeles; mientras que las demás métricas: *Spatial Information*, *Blurring*, *Ringing*, *Blockloss* y *Blocking*, se mantengan a la resolución original de 1920x1080 píxeles. Esta optimización reduce el tiempo medio de extracción de características de vídeo por imagen de 543.03 ms a 466.42 ms, lo que implica una disminución del coste computacional del 14.11 %. Además, el impacto en la precisión de la estimación de MOS es mínimo, con un error de solo 0.04, lo que valida la viabilidad de este enfoque.

Asimismo, se observa que el tiempo medio de extracción de características para las métricas *Temporal Information*, *Brightness* y *Contrast* procesadas en imágenes a una resolución de 320x180 píxeles es inferior a 1.3 ms por imagen (ver Tabla 4.2), y a 97.50 ms en medidas de tres segundos de duración. Estos resultados demuestran los beneficios de aplicar el procesamiento a baja resolución de estas tres métricas en todas las imágenes de la medida. El tiempo de 97.50 ms por medida se justifica siempre que se omita el procesamiento de estas tres métricas a resolución original en al menos 1.25 imágenes de media por medida. Esto se debe a que el procesamiento de *Temporal Information*, *Brightness* y *Contrast* a una resolución de 1920x1080 píxeles para una única imagen requiere 77.90 ms (ver Tabla 4.2).

En cuanto a la explotación de la redundancia temporal en secuencias de vídeo, se propone un enfoque basado en un muestreo temporal uniforme combinado con información sobre el tipo de imagen a nivel de codificación. Este enfoque, limitado a un máximo de cinco imágenes procesadas por medida de tres segundos de duración, garantiza el funcionamiento en tiempo real de la herramienta Video-MOS en el equipo de prueba. En este sentido, se propone el modo Q0-I-LIM5, el cual prioriza el procesamiento de la primera imagen de cada medida de tres segundos, así como de todas las imágenes tipo I, hasta alcanzar el umbral máximo de cinco imágenes utilizadas para la extracción de todas las características del vídeo. En caso de exceder dicho límite, las imágenes tipo I adicionales se descartan.

Los resultados obtenidos en el conjunto de secuencias de prueba indican que el modo Q0-I-LIM5 reduce el coste computacional medio a un 4.68 %, lo que implica un ahorro del 95.32 % respecto al modo de funcionamiento normal de la herramienta. Este coste computacional implica un procesamiento medio de únicamente 3.51 imágenes por medidas de tres segundos de duración, con una desviación estándar de 0.55. A pesar de la reducción considerable en el número de imágenes utilizadas para la extracción de todas las características del vídeo, el error en la estimación de MOS se mantiene en un valor de 0.14, lo que refleja una precisión aceptable en la evaluación de la calidad con este enfoque de optimización propuesto.

Finalmente, la combinación de la redundancia espacial y temporal da lugar a la propuesta del modo Q0-I-LIM5-RE, que integra ambas estrategias de optimización mencionadas anteriormente. El modo Q0-I-LIM5-RE contempla el procesamiento de las métricas *Temporal Information*, *Brightness* y *Contrast* a baja resolución en todas las imágenes de la medida, mientras que mantiene la extracción de características del resto de métricas a resolución original en un máximo de cinco imágenes, seleccionando la primera imagen de la medida y todas las imágenes tipo I disponibles. Por tanto, las métricas *Spatial Information*, *Blurring*, *Ringing*, *Blockloss* y *Blocking* se procesan con la resolución original de 1920x1080 píxeles, mientras que *Temporal Information*, *Brightness* y *Contrast* se procesan con una resolución reducida de 320x180 píxeles.

**Tabla 4.26:** Error de MOS vs. coste computacional en el enfoque final propuesto. 100 % de las secuencias de prueba.

| Modo | Imágenes (AVG) | Imágenes (STD) | Coste computacional ( %) | Error de MOS (MAE) |
|---|---|---|---|---|
| Q0-I-LIM5 | 3.51 | 0.55 | 4.68 | 0.14 |
| Q0-I-LIM5-RE | 3.51 | 0.55 | ↓ 4.68 | 0.09 |

**Tabla 4.27:** Error de MOS vs. coste computacional en el enfoque final propuesto. 17.36 % de las secuencias de prueba.

| Modo | Imágenes (AVG) | Imágenes (STD) | Coste computacional ( %) | Error de MOS (MAE) |
|---|---|---|---|---|
| Q0-I-LIM5 | 3.51 | 0.60 | 4.68 | 0.17 |
| Q0-I-LIM5-RE | 3.51 | 0.60 | ↓ 4.68 | 0.11 |

**Tabla 4.28:** Error en la extracción de características en el enfoque final propuesto.

| Métrica de evaluación | Error (MAE, en %) | |
|---|---|---|
| | **Q0-I-LIM5** | **Q0-I-LIM5-RE** |
| *Spatial Information* | 1.54 | 1.54 |
| *Temporal Information* | 4.74 | 3.29 |
| *Blurring* | 13.05 | 13.05 |
| *Brightness* | 2.09 | 0.12 |
| *Contrast* | 1.87 | 0.58 |
| *Ringing* | 3.64 | 3.64 |
| *Blockloss* | 1.95 | 1.95 |
| *Blocking* | 36.45 | 36.45 |
| **Vector de caract.** | **8.17** | **7.58** |

La Tabla 4.26 muestra la mejora obtenida al incorporar la redundancia espacial en el modo de redundancia temporal propuesto (modo Q0-I-LIM5 frente al modo Q0-I-LIM5-RE), para el conjunto total de secuencias de prueba. De igual manera, en la Tabla 4.27 se presentan los resultados obtenidos de esta comparativa utilizando el conjunto de datos correspondientes al 17.36 % de las secuencias más complejas consideradas en el estudio. Los datos de *Imágenes (AVG)*, *Imágenes (STD)* y *Coste computacional ( %)* hacen referencia a las imágenes en las que las métricas *Spatial Information*, *Blurring*, *Ringing*, *Blockloss* y *Blocking* se procesan a resolución original. En el modo Q0-I-LIM5-RE, que implica el procesamiento a baja resolución,

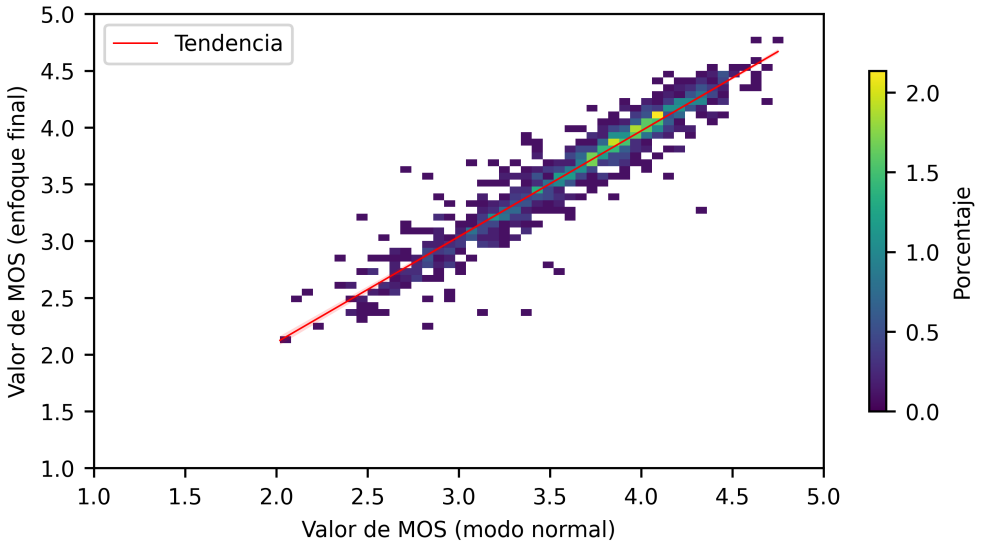


**Figura 4.27:** Histograma 2D de los valores MOS de las secuencias de prueba.

dado que el valor de *Imágenes (AVG)* es superior a 1.25 imágenes, el valor real de *Coste computacional ( %)* es inferior al presentado en la Tabla 4.26 y en la Tabla 4.27.

La mejora del modo Q0-I-LIM5-RE en comparación con el modo Q0-I-LIM5 es significativa en términos de error en la estimación de MOS, además de presentar un coste computacional medio inferior. Los errores de MOS obtenidos con el modo Q0-I-LIM5-RE quedan reducidos a 0.09 para el 100 % de las secuencias de prueba y a 0.11 para el conjunto del 17.36 % de secuencias más complejas.

La Tabla 4.28 presenta los valores de error de las diferentes métricas de vídeo y del vector de características para todo el conjunto de datos de prueba. El error en el vector de características es de 8.17 % en el modo Q0-I-LIM5, frente a un error de 7.58 % en el modo Q0-I-LIM5-RE. Esta mejora en el modo Q0-I-LIM5-RE se logra al procesar las métricas *Temporal Information*, *Brightness* y *Contrast* en todas las imágenes de la medida. No obstante, como se observa en la Tabla 4.28, el error obtenido en estas tres métricas para el modo Q0-I-LIM5-RE no es nulo, debido a que las métricas se procesan sobre imágenes con un tamaño de 320x180 píxeles.

Tras realizar todas las pruebas necesarias y analizar los resultados obtenidos, se concluye que el enfoque propuesto para optimizar el rendimiento computacional de la herramienta Video-MOS es el modo Q0-I-LIM5-RE. Esta propuesta garantiza el funcionamiento en tiempo real en todas las mediciones, independientemente del tipo y complejidad de la secuencia de vídeo, así como de la estructura del tamaño de GOP. Además, en el conjunto de secuencias de prueba, se alcanza un ahorro del coste computacional superior al 95.32 % en comparación con el modo de funcionamiento normal de la herramienta Video-MOS, con un error en la estimación del valor de MOS de 0.09.

La Figura 4.27 ilustra la comparativa entre los valores de MOS obtenidos en el modo de funcionamiento normal de la herramienta (que corresponden a los valores de *ground truth*), y los valores estimados mediante el enfoque final propuesto en esta tesis doctoral. Para

esta comparativa, se ajusta una línea de tendencia utilizando regresión lineal, para evaluar la relación entre los valores de *ground truth* y los estimados con el enfoque propuesto. Los resultados obtenidos muestran la precisión conseguida en la estimación de MOS con la solución de optimización propuesta.

# Capítulo 5

# Discusión

Este capítulo presenta una validación exhaustiva y detallada del enfoque final propuesto para la optimización del coste computacional en la solución Video-MOS, utilizando el modo Q0-I-LIM5-RE, descrito y analizado en profundidad en el Capítulo 4.

Para asegurar el correcto funcionamiento del modo propuesto en tiempo real en cualquier escenario posible, se han empleado seis contenidos de 24 horas de duración, correspondientes a seis de los canales HD más relevantes de la TDT pública en España: La1 HD, La2 HD, Antena3 HD, Cuatro HD, Telecinco HD y LaSexta HD. Las 144 horas de contenido audiovisual, junto con la diversidad tanto en el tipo de contenido (noticias, deportes, musicales, documentales, películas, series, etc.) como en los distintos radiodifusores y proveedores de contenidos, aseguran una validación completa del enfoque propuesto en esta tesis doctoral.

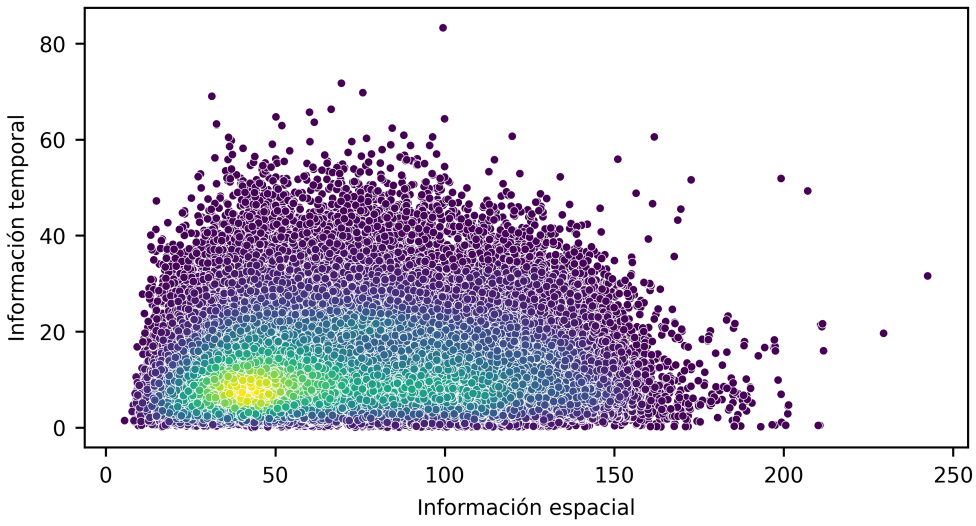


**Figura 5.1:** Diagrama SI-TI de las secuencias de validación.

El conjunto de datos de validación implica un total de más de 174000 medidas de vídeo de tres segundos de duración: 29416 medidas del canal La1 HD, 29416 medidas de La2 HD, 28794 de

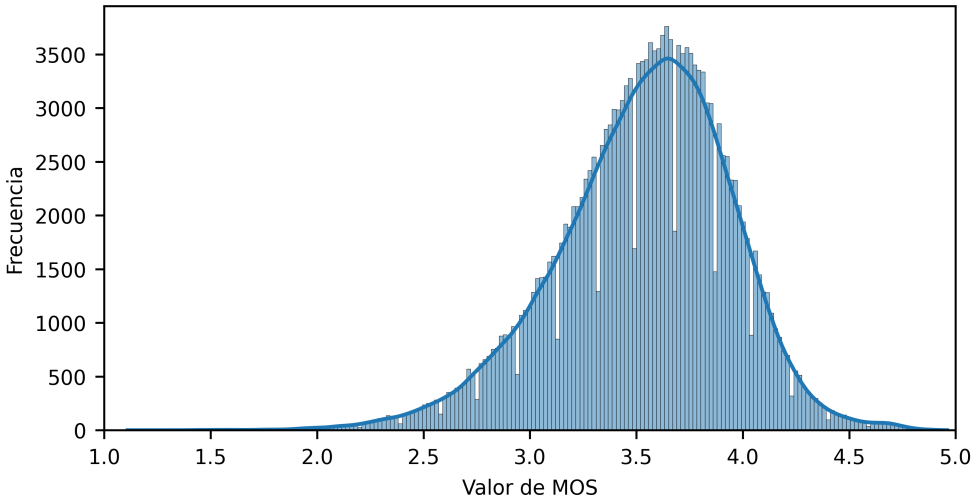


**Figura 5.2:** Histograma valor de MOS de las secuencias de validación.

Antena3 HD, 28833 de Cuatro HD, 28833 de Telecinco HD y 28794 de LaSexta HD. En la Figura 5.1 se muestra el diagrama SI-TI con todas las secuencias de vídeo pertenecientes al contenido de validación.

En la Figura 5.2 se muestra el histograma de los valores de MOS correspondientes a todas las secuencias de vídeo. El valor medio de MOS en el conjunto de validación es de 3.53, con un valor máximo de MOS de 4.96 y un valor mínimo de 1.11. Los resultados muestran que la mayoría de las secuencias de vídeo cumplen con los estándares de calidad requeridos para emisiones de contenidos HD en la TDT, según lo establecido en la norma EBU R132 [221]. Únicamente el 10.14 % de las secuencias de validación tienen un valor de MOS inferior a 3.

El análisis de la Guía Electrónica de Programación (EPG, *Electronic Program Guide*) permite determinar el tipo de contenido de los programas emitidos a través de los diferentes canales de televisión, siguiendo las directrices establecidas por la ETSI (*European Telecommunications Standards Institute*) en la especificación ETSI EN 300 468 [236], que define las normas relativas a la EPG en la TDT en España. Para llevar a cabo este análisis, se ha utilizado la herramienta de código abierto TSDuck [1], ampliamente reconocida en el ámbito audiovisual por su capacidad para analizar, procesar y generar transmisiones de video digital encapsuladas en TS (Transport Stream). La Figura 5.3 muestra un análisis de la distribución por tipo de contenido, basado en la clasificación de programas según la especificación ETSI EN 300 468 [236]. De los 52 programas de televisión analizados en las 144 horas de duración total, los datos porcentuales reflejan lo siguiente: 38.46 % de los programas pertenecen a la categoría *Show/Game show*, 21.15 % a la categoría *News/Current affairs*, 13.46 % a *Social/Political issues/Economics*, 9.62 % a *Undefined content*, 9.62 % a *Movie/Drama*, 5.77 % a *Education/Science/Factual topics* y 1.92 % a *Leisure hobbies*.

[1] https://tsduck.io/

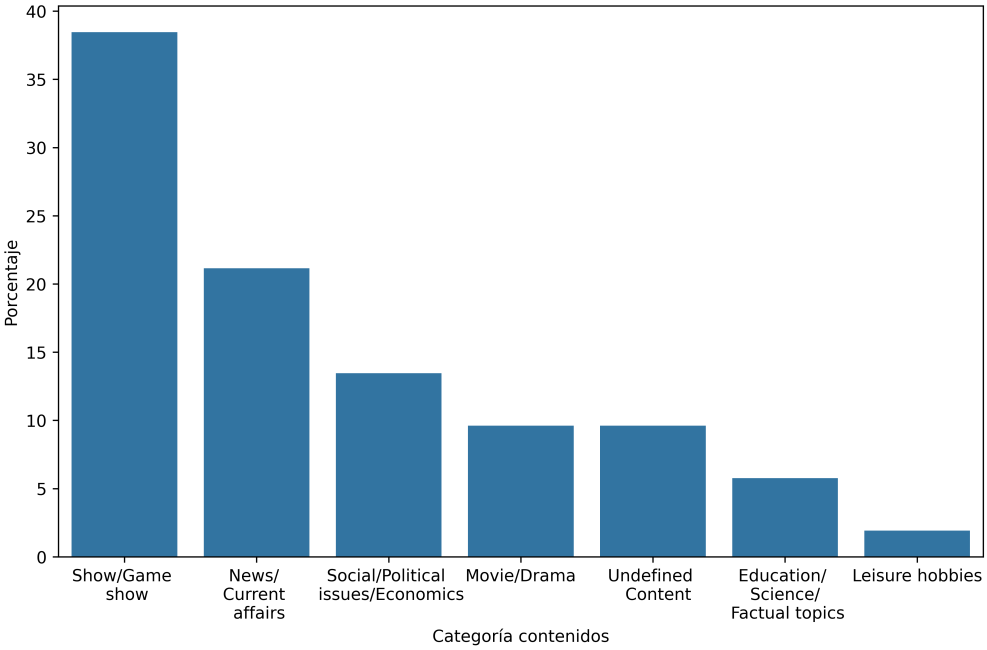


**Figura 5.3:** Tipo de contenido según la EPG de las secuencias de validación.

La Figura 5.4 y la Figura 5.5 presentan el análisis de los datos en relación con el número de imágenes tipo I por medida de tres segundos de duración, así como la distribución del tamaño de los GOP para cada uno de los seis canales de televisión utilizados en la validación.

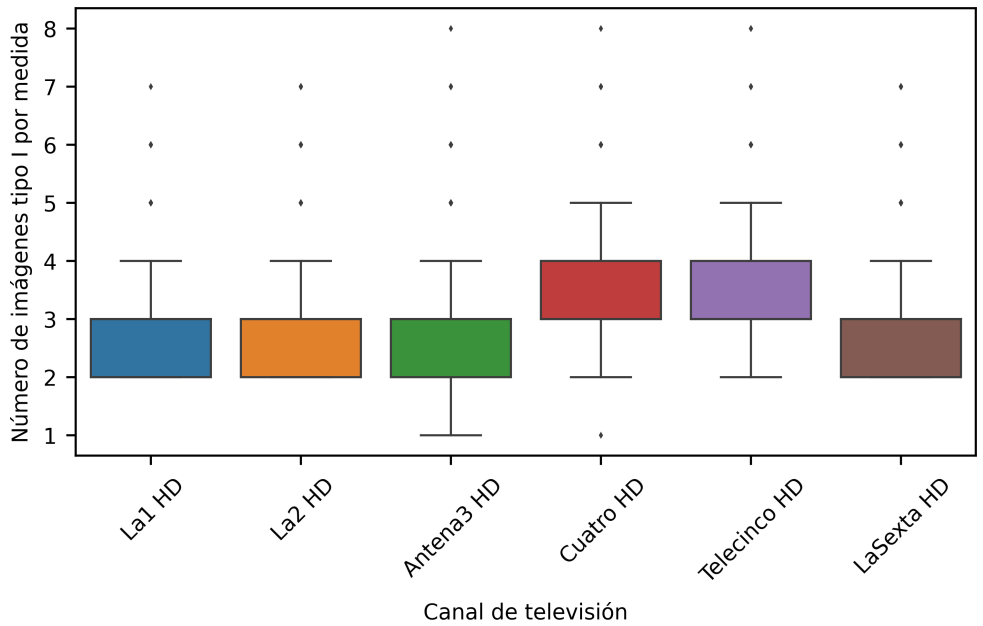


**Figura 5.4:** Número de imágenes tipo I por canal de televisión en las secuencias de validación.

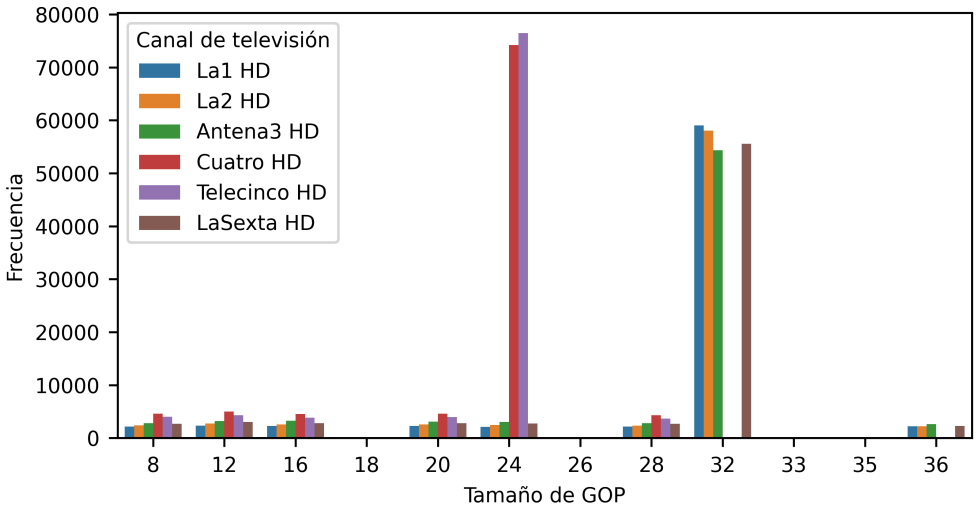


**Figura 5.5:** Tamaño de los GOP por canal de televisión en las secuencias de validación.

Los resultados obtenidos son consistentes con los de las secuencias de prueba, lo que valida el uso de imágenes tipo I en el enfoque final propuesto para la evaluación de calidad en tiempo real en contenidos HD de la TDT en España. La Tabla 5.1 muestra el análisis detallado de los tamaños de GOP (mayor, menor, promedio y más repetido) para cada canal de televisión.

**Tabla 5.1:** Tamaño de los GOP por canal de televisión en las secuencias de validación.

| Medida | La1 HD | La2 HD | Antena3 HD | Cuatro HD | Telecinco HD | LaSexta HD |
|---|---|---|---|---|---|---|
| Nº GOPs analizados | 74543 | 75331 | 75160 | 97246 | 96298 | 74571 |
| Tamaño GOP mayor | 44 | 36 | 62 | 43 | 47 | 52 |
| Tamaño GOP menor | 8 | 8 | 8 | 8 | 8 | 8 |
| Tamaño GOP medio | 29.60 | 29.29 | 28.73 | 22.23 | 22.45 | 28.96 |
| Tamaño de GOP más repetido | 32 | 32 | 32 | 24 | 24 | 32 |

Los tamaños de GOP más frecuentes son de 32 imágenes en los canales La1 HD, La2 HD, Antena3 HD y LaSexta HD, y de 24 imágenes en los canales Cuatro HD y Telecinco HD. El análisis de los datos revela un claro predominio de estos tamaños de GOP según el canal de televisión. En términos porcentuales con respecto al total de los GOPs por canal, el tamaño de GOP de 32 imágenes se repite en el 79 % de los GOPs de La1 HD, el 77 % de La2 HD, el 72 % de Antena3 HD y el 74 % de LaSexta HD. De manera similar, el tamaño de GOP de 24 imágenes se presenta en el 76 % de los GOPs de Cuatro HD y en el 79 % de Telecinco HD. Los tamaños de GOP de 32 y 24 imágenes, ambos con una estructura $IBBBP$, resultan en una media de 2.34 y 3.13 imágenes de tipo I por medida de tres segundos de vídeo.

El tamaño de GOP más pequeño, compuesto por 8 imágenes, está presente en los seis canales de televisión analizados. Además, otros tamaños de GOP que se repiten de manera significativa en todos los canales de televisión son 12, 16, 20, 28 y 36, como se puede observar

en la Figura 5.5. El tamaño máximo de GOP varía según el canal de televisión: 44 para el canal La1 HD, 36 para La2 HD, 62 para Antena3 HD, 43 para Cuatro HD, 47 para Telecinco HD y 52 para LaSexta HD. El tamaño medio de GOP es aproximadamente 29 imágenes en los canales La1 HD, La2 HD, Antena3 HD y LaSexta HD, mientras que en los canales Cuatro HD y Telecinco HD es de 22 imágenes. Un tamaño de GOP más pequeño, como ocurre en los canales Cuatro HD y Telecinco HD, implica un mayor número de imágenes tipo I en las medidas de vídeo, mientras que un tamaño de GOP mayor está asociado a un menor número de imágenes tipo I por medida (ver Figura 5.4).

En términos de coste computacional medio y error de MOS obtenido para el modo propuesto (Q0-I-LIM5-RE), la Tabla 5.2 presenta los datos tanto de manera individual para cada canal de televisión, como combinando todos los canales. Para los canales de Cuatro HD y Telecinco HD, que presentan tamaños de GOP más pequeños y, por tanto, un mayor número de imágenes tipo I por medida de tres segundos, se registran los valores más elevados de coste computacional medio, con 5.73 % y 5.69 %, respectivamente. En estos canales se procesan de media a resolución original, 4.30 y 4.27 imágenes del total de las 75 imágenes que componen las medidas de tres segundos. Este mayor coste computacional, sin superar el umbral de cinco imágenes que garantiza el funcionamiento en tiempo real, hace que se obtenga una estimación de calidad con el menor error de MOS para los diferentes canales de televisión: 0.09 en el canal Cuatro HD y 0.10 en Telecinco HD. En los otros cuatro canales restantes, el coste computacional medio oscila entre 4.66 % y 4.75 %, con errores de MOS que varían entre 0.11 y 0.14.

**Tabla 5.2:** Error de MOS vs. coste computacional en el enfoque final propuesto con las secuencias de validación.

| **Canal TV** | **Imágenes (AVG)** | **Imágenes (STD)** | **Coste computacional (%)** | **Error de MOS (MAE)** |
|---|---|---|---|---|
| La1 HD | 3.50 | 0.55 | ↓ 4.66 | 0.14 |
| La2 HD | 3.52 | 0.56 | ↓ 4.69 | 0.11 |
| Antena3 HD | 3.56 | 0.59 | ↓ 4.75 | 0.12 |
| Cuatro HD | 4.30 | 0.48 | ↓ 5.73 | 0.09 |
| Telecinco HD | 4.27 | 0.46 | ↓ 5.69 | 0.10 |
| LaSexta HD | 3.54 | 0.58 | ↓ 4.73 | 0.12 |
| Combinado (6) | 3.78 | 0.64 | ↓ 5.04 | 0.11 |

Para el conjunto total de datos de los seis canales de televisión, el coste computacional medio es inferior a 5.04 % y el error de MOS obtenido es de 0.11. Este enfoque representa un ahorro de coste computacional superior al 94.96 % en comparación con el modo de funcionamiento normal de la herramienta de Video-MOS, garantizando el tiempo real en el equipo de prueba. Además, mantiene una estimación precisa de la calidad, con un error de MOS inferior al valor máximo aceptado de 0.15.

La Tabla 5.3 muestra el error cometido en las métricas de vídeo y en el vector de características, al comparar la solución final propuesta con el modo de funcionamiento normal

**Tabla 5.3:** Error en la extracción de características en el enfoque final propuesto con las secuencias de validación.

| Métrica de evaluación | Error (MAE, en %) | | | | | | |
|---|---|---|---|---|---|---|---|
| | La1 HD | La2 HD | Antena3 HD | Cuatro HD | Telecinco HD | LaSexta HD | Combinado (6) |
| *Spatial Information* | 1.73 | 1.97 | 1.90 | 1.36 | 1.41 | 1.70 | 1.68 |
| *Temporal Information* | 4.73 | 1.13 | 3.82 | 2.62 | 3.10 | 5.92 | 3.55 |
| *Blurring* | 19.61 | 6.26 | 14.91 | 11.58 | 11.10 | 16.60 | 13.34 |
| *Brightness* | 0.19 | 0.09 | 0.13 | 0.31 | 0.15 | 0.16 | 0.17 |
| *Contrast* | 0.56 | 0.33 | 0.54 | 0.59 | 0.53 | 0.58 | 0.52 |
| *Ringing* | 6.67 | 3.92 | 6.30 | 4.41 | 4.50 | 5.83 | 5.27 |
| *Blockloss* | 4.47 | 2.22 | 4.08 | 3.46 | 2.92 | 5.01 | 3.69 |
| *Blocking* | 15.48 | 18.53 | 16.68 | 16.26 | 16.16 | 16.35 | 16.58 |
| **Vector de caract.** | **6.68** | **4.31** | **6.05** | **5.07** | **4.98** | **6.52** | **5.60** |

de la herramienta. Los valores de error obtenidos para las distintas métricas de vídeo son muy similares, independientemente del canal de televisión. Además, para los seis canales, las métricas de vídeo con los errores más altos se corresponden a las métricas de *Blurring* y *Blocking*.

En cuanto al vector de características, el error obtenido varía entre 4.31 % y 6.68 % para los seis canales de televisión, siendo un error de 5.60 % cuando se combinan todos los datos. Los canales Cuatro HD y Telecinco HD presentan errores en el vector de características más pequeños (5.07 % y 4.98 %, respectivamente) con respecto al resto de canales de televisión. Los valores de la Tabla 5.3 son comparables a los obtenidos con el conjunto de secuencias de vídeo de prueba, donde el error en el vector de características fue de 7.58 % con el modo final Q0-I-LIM5-RE.

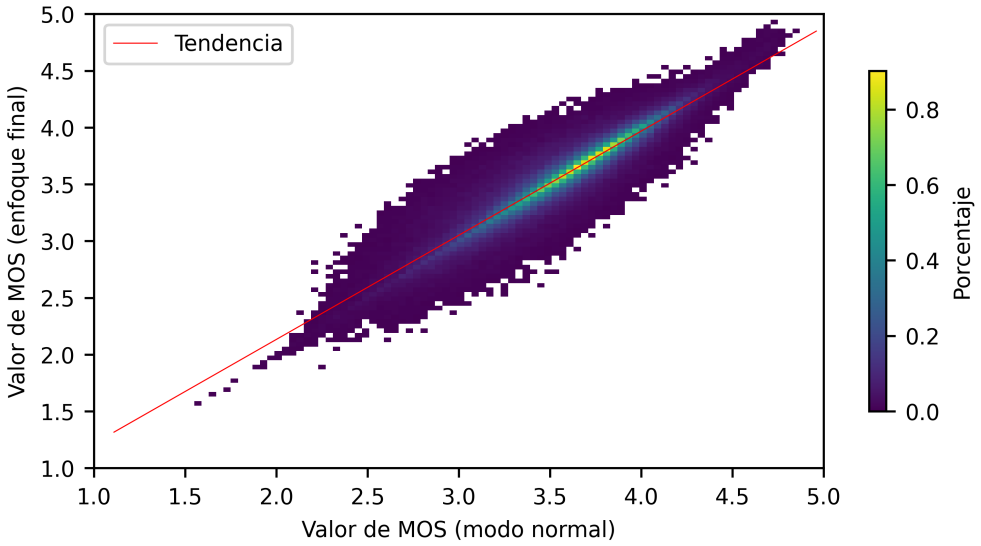


**Figura 5.6:** Histograma 2D de los valores MOS de las secuencias de validación.

Los resultados obtenidos en esta validación, realizada con más de 144 horas de contenidos variados emitidos en la TDT en España, demuestran la eficacia de la solución propuesta, la cual permite un ahorro significativo en el coste computacional de la herramienta Video-MOS sin comprometer la precisión en la estimación de calidad. La Figura 5.6 presenta la comparativa entre los valores de MOS obtenidos en el modo de funcionamiento normal de la herramienta Video-MOS (valores *ground truth*) y los valores estimados mediante el enfoque propuesto. Al igual que con el conjunto de datos de prueba, los resultados evidencian la precisión alcanzada con la solución de optimización propuesta. Además, en la representación gráfica, se ha ajustado una línea de tendencia mediante regresión lineal para evaluar la relación entre los dos conjuntos de valores.

# Capítulo 6

# Conclusiones

En este capítulo se presentan las conclusiones extraídas de todo el trabajo desarrollado en esta tesis doctoral, resaltando las contribuciones más relevantes y el impacto de los resultados obtenidos. Asimismo, se analizan las principales limitaciones del enfoque propuesto y se plantean posibles líneas de investigación futura que permitirían ampliar y mejorar los alcances de este estudio realizado. Estas direcciones de trabajo buscan optimizar aún más el rendimiento computacional de la herramienta Video-MOS, mejorar la estimación de la calidad, explorar nuevas metodologías y adaptar la solución a escenarios mucho más complejos y exigentes.

## 6.1 Conclusiones

La evaluación automática de la calidad percibida en contenidos audiovisuales ha adquirido una relevancia creciente en los últimos años, impulsada tanto por la evolución de la TDT en países como España como por el constante incremento del tráfico de vídeo sobre redes IP. En este contexto, competir en calidad se ha convertido en un factor clave para que los distintos actores y plataformas logren diferenciarse y fidelizar a sus usuarios en un entorno cada vez más saturado y exigente. Este continuo desarrollo de herramientas de evaluación automática responde a la necesidad de controlar y garantizar la calidad ofrecida ante la creciente variedad de contenidos, los servicios de *streaming* y la diversidad de terminales disponibles, que han transformado y diversificado profundamente los modelos de consumo audiovisual. Todo ello ha generado la necesidad de desarrollar herramientas específicas y precisas para la medición objetiva de la calidad del vídeo en estos entornos dinámicos y en constante evolución.

El estudio de la calidad visual ha sido objeto de investigación durante décadas, dando lugar a una amplia variedad de enfoques y modelos presentes en la literatura actual. Si bien la evaluación subjetiva es considerada el método más fiable para medir la calidad percibida, su aplicación en escenarios de tiempo real resulta inviable debido a la complejidad de las metodologías empleadas y a la necesidad de contar con un número significativo de observadores reales. En este contexto, ha surgido un creciente interés por el desarrollo de métricas objetivas de evaluación de vídeo, en particular las métricas NR-VQA, las cuales permiten estimar la calidad visual sin necesidad de disponer de una señal de referencia. Estas métricas han demostrado ser una alternativa eficiente y viable para la evaluación de calidad en aplicaciones

en tiempo real, facilitando la automatización del proceso y reduciendo la carga computacional asociada a métodos tradicionales.

Esta tesis doctoral ha llevado a cabo un análisis exhaustivo de los distintos modelos de evaluación objetiva de calidad audiovisual disponibles en la literatura, abarcando desde métodos convencionales y tradicionales hasta los enfoques más recientes basados en IA y en aprendizaje automático. En particular, se ha puesto especial énfasis en los modelos NR-VQA, los cuales enfrentan numerosos desafíos debido a la constante evolución de los formatos audiovisuales. Factores como el aumento de las resoluciones del vídeo, las mayores frecuencias de refresco y la aparición de tecnologías emergentes como el HDR y el WCG (*Wide Color Gamut*), requieren una continua adaptación de estos modelos para mantener su precisión y fiabilidad. Además, la necesidad de optimizar el coste computacional de estas métricas se vuelve fundamental, dado que deben procesar grandes volúmenes de datos de vídeo en entornos cada vez más exigentes y con restricciones de tiempo real.

En esta investigación se han desarrollado y evaluado diferentes estrategias orientadas a la reducción del coste computacional en el procesamiento de vídeo para la estimación de la QoE. Aunque dichas estrategias han sido implementadas y validadas en la herramienta Video-MOS, los conocimientos derivados de este estudio son aplicables a otros modelos y enfoques de evaluación de calidad audiovisual. Siguiendo la lógica de los estándares de codificación de vídeo, que optimizan la transmisión de información mediante la explotación de la redundancia espacial y temporal de los vídeos, los enfoques propuestos en esta tesis doctoral se fundamentan en la reducción del tamaño de las imágenes procesadas y en la limitación del número de imágenes analizadas dentro de una secuencia de vídeo. Las pruebas experimentales realizadas para cada uno de estos enfoques han permitido evaluar su eficacia en términos de reducción del coste computacional, impacto en la precisión de las métricas de vídeo, errores en la extracción de las características de la imagen y efecto sobre la estimación del valor de MOS. Los resultados obtenidos han sido contrastados sobre un amplio conjunto de secuencias de prueba, procedentes de emisiones de TDT en España. Dicho conjunto está compuesto por 1123 medidas de vídeo de tres segundos de duración, con una resolución de imagen de 1920x1080 píxeles y una frecuencia de refresco de 25 imágenes por segundo.

El análisis de la redundancia espacial en secuencias de vídeo ha sido abordado en esta investigación como una estrategia para la reducción del coste computacional asociado al procesamiento del vídeo. Con este propósito, se han explorado distintos enfoques, incluyendo la selección específica de regiones dentro de la imagen, la identificación de áreas de interés mediante modelos de detección de saliencia y cambios de resolución a tamaños de imagen más pequeños. En particular, la utilización de resoluciones reducidas, tales como 1280x720, 960x540, 640x360, 480x270 y 320x180, en contraste con la resolución original de 1920x1080 píxeles, ha demostrado ser una solución efectiva para minimizar el coste computacional. Esta reducción de tamaño conlleva una disminución notable en el tiempo de procesamiento de las imágenes, así como en la extracción de características y en el cálculo de las métricas de vídeo, lo que permite optimizar el desempeño de la herramienta Video-MOS.

El enfoque basado en la selección de la región central de la imagen presenta una limitación significativa, ya que omite información relevante contenida en las áreas periféricas, como los elementos gráficos superpuestos en las secuencias de vídeo. Los resultados experimentales

indican que la reducción del tamaño de la región central conlleva un incremento en el error de estimación del valor de MOS. En particular, al disminuir la región analizada de 1280x720 a 320x180 píxeles, el error en la estimación de MOS se incrementa de 0.35 a 0.54, lo que evidencia una pérdida considerable de precisión en la evaluación de calidad. Por otro lado, la estrategia basada en la selección de regiones específicas de la imagen demostró ser más efectiva, alcanzando un error de MOS de 0.34 al emplear cuatro de las nueve regiones en las que se dividía la imagen. La utilización de métodos de detección de saliencia espacio-temporal para identificar automáticamente las áreas de interés no resultó ser una estrategia viable. A pesar de que esta técnica logró reducir el error de MOS a 0.29, el elevado coste computacional asociado al cálculo del mapa de saliencia limitó su aplicabilidad en entornos de tiempo real.

Finalmente, se exploró la estrategia de cambio de resolución de las imágenes a tamaños más pequeños. El proceso de cambio de resolución implica un submuestreo de píxeles, lo cual conlleva una pérdida de detalles y altera la estructura de la imagen, afectando negativamente la precisión en la estimación del valor final de MOS. Durante las pruebas, se utilizaron diversos métodos de interpolación para realizar los cambios de resolución, y se observó que la interpolación cúbica ofrecía el mejor equilibrio entre el tiempo de procesamiento y la calidad de la imagen resultante. En cuanto a los resultados, el error de MOS al reducir la resolución a 1280x720 fue de 0.53, y aumentó a 0.82 al reducir la resolución a 320x180 píxeles. Sin embargo, al procesar únicamente las métricas de vídeo *Temporal Information*, *Brightness* y *Contrast* a baja resolución de 320x180, y mantener el procesamiento de las demás métricas a la resolución original de 1920x1080 píxeles, se lograron resultados prometedores. El error de MOS se redujo a 0.04, mientras que se obtuvo un ahorro computacional del 14.13 % en la extracción de características de la imagen. Este enfoque, al limitarse a estas tres métricas a baja resolución, demostró ser una solución eficaz que combina reducción de costes computacionales con una buena precisión en la estimación de la calidad.

La redundancia temporal en los vídeos se origina por la similitud de información entre las imágenes consecutivas en una secuencia de vídeo. Los enfoques basados en muestreo temporal uniforme han demostrado que, al reducir el número de imágenes procesadas para la extracción de características, se consigue una disminución significativa del coste computacional, aunque esto conlleva un incremento en el error en la estimación del valor de MOS. En este estudio se analizaron varios modos de muestreo temporal uniforme, entre los que se incluyen MOD2, MOD5, MOD10, MOD15, MOD20, MOD25, MOD38 y Q0. El modo Q0, que solo procesa la primera imagen de la secuencia, representa el caso extremo en términos de reducción de procesamiento. Por otro lado, el modo MOD15, que selecciona una imagen de cada quince, proporciona un buen compromiso entre eficiencia y precisión. El modo MOD15 permitió el funcionamiento en tiempo real de la solución Video-MOS en el equipo de prueba, independientemente de la complejidad o tipo de contenido de la secuencia de vídeo, procesando una de cada quince imágenes. Los resultados obtenidos con este modo muestran un coste computacional promedio del 6.68 %, lo que representa un ahorro computacional del 93.32 % respecto al funcionamiento normal de la herramienta Video-MOS. Además, con el modo MOD15 se procesaron 5.00 imágenes de media por medida de tres segundos de duración, con un error en la estimación del valor de MOS de 0.14.

Además de los enfoques por muestreo temporal uniforme, también se implementó el uso

de la métrica SSIM para detectar cambios significativos entre imágenes consecutivas, con el objetivo de procesar aquellas imágenes que mostraban variaciones notables en el vídeo. El umbral de SSIM utilizado tuvo un impacto directo en la efectividad de los enfoques, y se identificó que un valor de SSIM de 0.3 representaba el equilibrio ideal entre el coste computacional y la precisión en la estimación de la calidad. El modo más eficiente en cuanto al número de imágenes procesadas para garantizar el funcionamiento en tiempo real fue el modo MOD25-SSIM03-LIM5. Este modo alcanzó un coste computacional promedio del 4.94 %, lo que implica un ahorro del 95.06 % respecto al modo de funcionamiento normal, procesando un promedio de 3.70 imágenes por medida de tres segundos. Además, el error de estimación del valor de MOS fue de 0.16. Sin embargo, el alto coste computacional asociado con el cálculo de la métrica SSIM en todas las imágenes de la medida y los resultados no tan prometedores en términos de precisión final descartaron este enfoque.

En esta investigación también se evaluaron enfoques de muestreo temporal uniforme combinados con el uso de imágenes con nivel de codificación tipo I, las cuales tienen la particularidad de que se codifican sin recurrir a predicciones basadas en otras imágenes y sirven como referencia para las demás. Las imágenes tipo I suelen corresponder a puntos en la secuencia que implican un cambio de escena significativo. El uso de estas imágenes para la extracción de características resultó ser particularmente eficaz, ya que permitió reducir el número de imágenes procesadas mediante muestreos temporales más amplios, sin comprometer la precisión en la estimación de la calidad. El modo Q0-I-LIM5, que también limita la cantidad de imágenes procesadas para garantizar el funcionamiento en tiempo real, alcanzó un coste computacional medio de 4.68 %, lo que representa un ahorro del 95.32 % en comparación con el modo normal. Este enfoque procesó un promedio de 3.51 imágenes por medida de tres segundos de duración y obtuvo un error de estimación del MOS de 0.14.

Para optimizar la eficiencia del coste computacional, se integraron las técnicas más prometedoras previamente exploradas, aprovechando tanto la redundancia espacial como temporal presente en las secuencias de vídeo. El enfoque final propuesto se basa en la extracción de características de vídeo, concretamente las métricas de *Temporal Information*, *Brightness* y *Contrast*, procesadas a baja resolución de 320x180 píxeles en todas las imágenes de la medida. Las métricas restantes de la herramienta Video-MOS se procesan a la resolución original de 1920x1080 píxeles, limitándose a la primera imagen de la secuencia de vídeo y a todas las imágenes tipo I de la medida. Además, el modo Q0-I-LIM5-RE también impone un límite de cinco en el número máximo de imágenes procesadas para garantizar el funcionamiento en tiempo real en el equipo de prueba. Los resultados obtenidos al aplicar este enfoque sobre el conjunto de prueba, compuesto por las 1123 medidas de vídeo, mostraron un coste computacional promedio inferior al 4.68 %, lo que supone un ahorro medio superior al 95.32 %. El promedio de imágenes procesadas por medida de tres segundos fue de 3.51, y el error en la estimación del valor de MOS fue de 0.09.

La estrategia más eficiente para mejorar el rendimiento computacional de la herramienta Video-MOS consistió en la combinación de la extracción de características a baja resolución en métricas de vídeo específicas, junto con el uso de muestreo temporal uniforme y la inclusión de imágenes tipo I. La validación exhaustiva del enfoque propuesto, realizada con más de 144 horas de vídeo procedente de seis de los principales canales HD de la TDT en España (La1

HD, La2 HD, Antena3 HD, Cuatro HD, Telecinco HD y LaSexta HD), confirma la validez de la solución. Este éxito se debe en gran medida a los tamaños de GOP típicamente utilizados en la codificación de vídeo H.264/AVC para contenido HD en la TDT en España, donde los más comunes son de 24 o 32 imágenes. Estos tamaños de GOP implican un promedio de 3.13 o 2.34 imágenes tipo I, respectivamente, por medida de tres segundos de vídeo.

Para todo el conjunto de datos validación, compuesto por más de 174000 medidas de tres segundos de duración, la solución final propuesta ofreció un coste computacional promedio inferior a 5.04 %, lo que representa un ahorro computacional de más del 94.96 % en comparación con el funcionamiento normal de la solución Video-MOS. En términos de procesamiento, se emplearon en promedio 3.78 imágenes por medida de tres segundos procesadas a resolución original, con un error de estimación de MOS de 0.11. Estos resultados resaltan la efectividad del enfoque propuesto en términos de optimización computacional y precisión en la estimación de la calidad del vídeo.

## 6.2 Líneas de trabajo futuro

En esta investigación se han obtenido resultados prometedores que permiten alcanzar ahorros significativos en el coste computacional de la herramienta Video-MOS, sin comprometer la precisión en la estimación del valor de MOS. A pesar de que las pruebas y los resultados fueron obtenidos utilizando la herramienta de desarrollo de Video-MOS y en un equipo de prueba con características técnicas específicas, se espera que los conocimientos adquiridos sean escalables y puedan aplicarse de manera efectiva a la versión comercial de Video-MOS, así como a sistemas y equipos con diferentes especificaciones técnicas.

Una de las principales áreas de investigación identificadas en esta tesis doctoral es determinar la cantidad máxima de imágenes que se pueden procesar para garantizar el funcionamiento en tiempo real de la herramienta Video-MOS, considerando las especificaciones técnicas del equipo en el que se despliegue la solución. De igual manera que se estableció un límite de cinco imágenes máximas a procesar para asegurar el tiempo real en el equipo de prueba según un conjunto de pruebas realizadas, será necesario desarrollar un módulo o algoritmo que, basado en el rendimiento específico del equipo, calcule la cantidad máxima de imágenes que podrían procesarse para realizar la extracción completa de las características del vídeo sin comprometer el funcionamiento en tiempo real de la herramienta.

Cualquier nueva métrica de vídeo incorporada en la solución Video-MOS y utilizada para la estimación de calidad deberá someterse a un estudio detallado para evaluar su impacto en el coste computacional final de la herramienta. Este análisis deberá determinar las estrategias de optimización a aplicar en función de la naturaleza de la métrica, diferenciando si se basa en información relacionada con la estructura de la imagen o en los niveles de los píxeles de la imagen.

Además, otras áreas de investigación deberán enfocarse en analizar cómo los nuevos formatos audiovisuales afectan al enfoque final propuesto. Esto incluye la evaluación de resoluciones de vídeo como 4K y 8K, así como las altas tasas de refresco de la imagen. Será igualmente crucial investigar el impacto de los nuevos estándares de codificación de vídeo como H.265/HEVC o

H.266/VVC, así como los posibles cambios en los tamaños de GOP en las emisiones de TDT en España y en otros países. Adicionalmente, se sugiere explorar la robustez de la solución propuesta en la transmisión de contenidos audiovisuales a través de Internet, teniendo en cuenta los estándares y protocolos comúnmente empleados en el tráfico de vídeo por IP, así como en servicios de *streaming* adaptativo, ya que estos aspectos podrían influir tanto en la eficiencia como en la precisión del enfoque de optimización desarrollado en esta tesis doctoral.

Asimismo, resulta relevante analizar cómo integrar la solución propuesta en los sistemas de gestión de redes móviles y fijas, de modo que los operadores de red puedan aplicar estrategias de optimización de los recursos disponibles, con el objetivo de lograr una gestión y distribución más eficientes de los contenidos audiovisuales.

# Referencias


[1] Observatorio Nacional de las Telecomunicaciones y de la SI, *Estudio de uso y actitudes de consumo de contenidos digitales*, Ministerio de Energía, Turismo y Agenda Digital, Accessed: Feb. 24, 2024, 2017.

[2] V. Chaudhari y M. Damle, "Impact of Covid Pandemic on OTT Streaming Services", en *2023 Somaiya International Conference on Technology and Information Management (SICTIM)*, 2023, págs. 59-63. DOI: 10.1109/SICTIM56495.2023.10105083.

[3] Cisco, *Cisco Visual Networking Index: Forecast and Trends, 2017–2022*, Cisco public white paper, 2019.

[4] Sandvine Phenomena, *Growing app complexity: Paving the way for digital lifestyles and immersive experiences*, The global Internet phenomena report, 2022.

[5] Z. Zhu, W. Sun, J. Jia et al., *Subjective and Objective Quality-of-Experience Evaluation Study for Live Video Streaming*, 2024. arXiv: 2409.17596 [cs.MM].

[6] N. Eswara, K. Manasa, A. Kommineni et al., "A Continuous QoE Evaluation Framework for Video Streaming Over HTTP", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, n.º 11, págs. 3236-3250, 2018. DOI: 10.1109/TCSVT.2017.2742601.

[7] P. Chen, L. Li, Y. Huang, F. Tan y W. Chen, "QoE Evaluation for Live Broadcasting Video", en *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, págs. 454-458. DOI: 10.1109/ICIP.2019.8802978.

[8] International Telecommunication Union, *ITU Recommendation P.10: Image parameter values for digital static television: Recommendation ITU-T P.10 (11/2017)*, ITU Recommendation, Accessed: Nov. 17, 2023, 2017.

[9] Z. Akhtar, K. Siddique, A. Rattani, S. L. Lutfi y T. H. Falk, "Why is Multimedia Quality of Experience Assessment a Challenging Problem?", *IEEE Access*, vol. 7, págs. 117 897-117 915, 2019. DOI: 10.1109/ACCESS.2019.2936470.

[10] N. Somraj, M. S. Kashi, S. Arun y R. Soundararajan, "Understanding the perceived quality of video predictions", *Signal Processing: Image Communication*, vol. 102, pág. 116 626, 2022. DOI: https://doi.org/10.1016/j.image.2021.116626.

[11] J. Joskowicz y R. Sotelo, "Modelo de Estimación de Calidad de Video: Video Quality Experts Groups", Memoria de trabajos de difusión científica y técnica, inf. téc. 10, 2012, Accessed: Feb. 24, 2024.

[12] "Report: The Nielsen Total Audience Report: February 2020", Nielsen, feb. de 2020.

[13] J. Song, F. Yang, Y. Zhou, S. Wan y H. R. Wu, "QoE Evaluation of Multimedia Services Based on Audiovisual Quality and User Interest", *IEEE Transactions on Multimedia*, vol. 18, n.º 3, págs. 444-457, 2016. DOI: 10.1109/TMM.2016.2520090.

[14] U. Engelke, D. P. Darcy, G. H. Mulliken et al., "Psychophysiology-Based QoE Assessment: A Survey", *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, n.º 1, págs. 6-21, 2017. DOI: 10.1109/JSTSP.2016.2609843.

[15] C. W. Chen, P. Chatzimisios, T. Dagiuklas y L. Atzori, *Multimedia quality of experience (QoE): current status and future requirements*. John Wiley & Sons, 2015.

[16] T. Yamazaki, "Quality of experience (QoE) studies: Present state and future prospect", *IEICE Transactions on Communications*, vol. 104, n.º 7, págs. 716-724, 2021.

[17] G. Kougioumtzidis, V. Poulkov, Z. D. Zaharis y P. I. Lazaridis, "A Survey on Multimedia Services QoE Assessment and Machine Learning-Based Prediction", *IEEE Access*, vol. 10, págs. 19 507-19 538, 2022. DOI: 10.1109/ACCESS.2022.3149592.

[18] D. Li, T. Jiang y M. Jiang, "Recent Advances and Challenges in Video Quality Assessment", vol. 17, págs. 3-11, mar. de 2019. DOI: 10.12142/ZTECOM.201901002.

[19] M. Smirnov, A. Gushchin, A. Antsiferova et al., *AIM 2024 Challenge on Compressed Video Quality Assessment: Methods and Results*, 2024. arXiv: 2408.11982 [eess.IV].

[20] M. Cheon y J.-S. Lee, "Subjective and Objective Quality Assessment of Compressed 4K UHD Videos for Immersive Experience", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, n.º 7, págs. 1467-1480, 2018. DOI: 10.1109/TCSVT.2017.2683504.

[21] C. Bonnineau, W. Hamidouche, J. Fournier, N. Sidaty, J.-F. Travers y O. Déforges, "Perceptual Quality Assessment of HEVC and VVC Standards for 8K Video", *IEEE Transactions on Broadcasting*, vol. 68, n.º 1, págs. 246-253, 2022. DOI: 10.1109/TBC.2022.3140710.

[22] Y. Li, S. Meng, X. Zhang et al., "User-Generated Video Quality Assessment: A Subjective and Objective Study", *IEEE Transactions on Multimedia*, vol. 25, págs. 154-166, 2023. DOI: 10.1109/TMM.2021.3122347.

[23] Y. Sugito y M. Bertalmío, "PERFORMANCE EVALUATION OF OBJECTIVE QUALITY METRICS ON HLG-BASED HDR IMAGE CODING", en *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2018, págs. 96-100. DOI: 10.1109/GlobalSIP.2018.8646673.

[24] I. P. Gunawan, O. Cloramidina, S. B. Syafa'ah, R. H. Febriani, G. P. Kuntarto y B. I. Santoso, "A review on high dynamic range (HDR) image quality assessment", *International Journal on Smart Sensing and Intelligent Systems*, vol. 14, n.º 1, págs. 1-17, 2021. DOI: doi:10.21307/ijssis-2021-010.

[25] M. Xu, C. Li, Z. Chen, Z. Wang y Z. Guan, "Assessing Visual Quality of Omnidirectional Videos", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, n.º 12, págs. 3516-3530, 2019. DOI: 10.1109/TCSVT.2018.2886277.

[26] M. Xu, C. Li, S. Zhang y P. L. Callet, "State-of-the-Art in 360° Video/Image Processing: Perception, Assessment and Compression", *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, n.º 1, págs. 5-26, 2020. DOI: 10.1109/JSTSP.2020.2966864.

[27] Y. Jin, M. Chen, T. Goodall, A. Patney y A. C. Bovik, "Subjective and Objective Quality Assessment of 2D and 3D Foveated Video Compression in Virtual Reality",

*IEEE Transactions on Image Processing*, vol. 30, págs. 5905-5919, 2021. DOI: 10.1109/TIP.2021.3087322.

[28] M. S. Anwar, J. Wang, W. Khan, A. Ullah, S. Ahmad y Z. Fei, "Subjective QoE of 360-Degree Virtual Reality Videos and Machine Learning Predictions", *IEEE Access*, vol. 8, págs. 148 084-148 099, 2020. DOI: 10.1109/ACCESS.2020.3015556.

[29] N. Barman, S. Zadtootaghaj, S. Schmidt, M. G. Martini y S. Möller, "An objective and subjective quality assessment study of passive gaming video streaming", *International Journal of Network Management*, vol. 30, n.º 3, e2054, 2020, e2054 nem.2054. DOI: https://doi.org/10.1002/nem.2054. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nem.2054.

[30] Y. Gao, Y. Cao, T. Kou et al., "VDPVE: VQA Dataset for Perceptual Video Enhancement", en *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023, págs. 1474-1483. DOI: 10.1109/CVPRW59228.2023.00152.

[31] X. Jiang, H. Yao, S. Zhang, X. Lu y W. Zeng, "Night video enhancement using improved dark channel prior", en *2013 IEEE International Conference on Image Processing*, 2013, págs. 553-557. DOI: 10.1109/ICIP.2013.6738114.

[32] P. Da, G. Song, P. Shi y H. Zhang, "Perceptual quality assessment of nighttime video", *Displays*, vol. 70, pág. 102 092, 2021. DOI: https://doi.org/10.1016/j.displa.2021.102092.

[33] M. Nilsson, "Ultra High Definition Video Formats and Standardisation", *BT Media and Broadcast Research Paper*, 2015.

[34] ITU-R, *Parameter values for ultra-high definition television systems for production and international programme exchange, document Rec. BT.2020*, ITU-R Document, Accessed: Nov. 17, 2023, 2015.

[35] ITU-R, *Parameter values for the HDTV standards for production and international programme exchange, document Rec. BT.709*, ITU-R Document, Accessed: Nov. 17, 2023, 2015.

[36] International Energy Agency, "Electricity 2024: Analysis and forecast to 2026", International Energy Agency, 2024, Accessed: 2025-04-09.

[37] S. Afzal, N. Mehran, Z. A. Ourimi et al., *A Survey on Energy Consumption and Environmental Impact of Video Streaming*, 2024. arXiv: 2401.09854 [cs.MM].

[38] H. Bruck. "The Energy Cost of Social Media". Texas Public Policy Foundation, published July 16, 2024. Accessed April 8, 2025. (2024), dirección: https://www.texaspolicy.com/the-energy-cost-of-social-media/.

[39] R. Madlener, S. Sheykhha y W. Briglauer, "The electricity- and CO2-saving potentials offered by regulation of European video-streaming services", *Energy Policy*, vol. 161, pág. 112 716, 2022. DOI: https://doi.org/10.1016/j.enpol.2021.112716.

[40] G. Bingöl, A. Floris, S. Porcu, C. Timmerer y L. Atzori, "Are Quality and Sustainability Reconcilable? A Subjective Study on Video QoE, Luminance and Resolution", en *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*, 2023, págs. 19-24. DOI: 10.1109/QoMEX58391.2023.10178513.

[41] C. Herglotz, W. Robitza, A. Raake, T. Hossfeld y A. Kaup, *Power Reduction Opportunities on End-User Devices in Quality-Steady Video Streaming*, 2023. arXiv: 2305.15117 [eess.IV].

[42] G. Bingöl, "Sustainability vs. Quality of Experience: Striking the Right Balance for Video Streaming", *SIGMultimedia Rec.*, vol. 15, n.º 2, dic. de 2024. DOI: 10.1145/3708973.3708975.

[43] T. L. Project, "Quantitative study of the GHG emissions of delivering TV content: Final Report – Version 1.2", The LoCaT Project, inf. téc., 2021, Accessed: Mar. 18, 2024.

[44] G. Bingöl, S. Porcu, A. Floris y L. Atzori, "An Analysis of the Trade-Off Between Sustainability and Quality of Experience for Video Streaming", *2023 IEEE International Conference on Communications Workshops (ICC Workshops)*, págs. 1600-1605, 2023.

[45] T. Hoßfeld, M. Varela, L. Skorin-Kapov y P. E. Heegaard, "A Greener Experience: Trade-Offs between QoE and CO2 Emissions in Today's and 6G Networks", *IEEE Communications Magazine*, vol. 61, n.º 9, págs. 178-184, 2023. DOI: 10.1109/MCOM.006.2200490.

[46] D. Y. Lee, S. Paul, C. G. Bampis et al., "A Subjective and Objective Study of Space-Time Subsampled Video Quality", *IEEE Transactions on Image Processing*, vol. 31, págs. 934-948, 2022. DOI: 10.1109/TIP.2021.3137658.

[47] J. Y. Lin, R. Song, C.-H. Wu, T. Liu, H. Wang y C.-C. J. Kuo, "MCL-V: A streaming video quality assessment database", *Journal of Visual Communication and Image Representation*, vol. 30, págs. 1-9, 2015. DOI: https://doi.org/10.1016/j.jvcir.2015.02.012.

[48] A. Mackin, M. Afonso, F. Zhang y D. Bull, "A Study of Subjective Video Quality at Various Spatial Resolutions", en *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, págs. 2830-2834. DOI: 10.1109/ICIP.2018.8451225.

[49] Q. Huang, S. Y. Jeong, S. Yang et al., "Perceptual Quality Driven Frame-Rate Selection (PQD-FRS) for High-Frame-Rate Video", *IEEE Transactions on Broadcasting*, vol. 62, n.º 3, págs. 640-653, 2016. DOI: 10.1109/TBC.2016.2570022.

[50] A. V. Katsenou, D. Ma y D. R. Bull, "Perceptually-Aligned Frame Rate Selection Using Spatio-Temporal Features", en *2018 Picture Coding Symposium (PCS)*, 2018, págs. 288-292. DOI: 10.1109/PCS.2018.8456274.

[51] R. R. Ramachandra Rao, S. Göring, W. Robitza, B. Feiten y A. Raake, "AVT-VQDB-UHD-1: A Large Scale Video Quality Database for UHD-1", en *2019 IEEE International Symposium on Multimedia (ISM)*, 2019, págs. 17-177. DOI: 10.1109/ISM46123.2019.00012.

[52] Y. Wang, Z. Chen, H. Jiang, S. Song, Y. Han y G. Huang, "Adaptive Focus for Efficient Video Recognition", en *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, págs. 16 229-16 238. DOI: 10.1109/ICCV48922.2021.01594.

[53] Z. Gao, G. Lu y P. Yan, "Key-frame selection for video summarization: an approach of multidimensional time series analysis", *Multidimensional Systems and Signal Processing*, vol. 29, págs. 1485-1505, 2018. DOI: 10.1007/s11045-017-0513-9.

[54] J. Karotte y E. Sarma, "An evaluation of the effect of image down-sampling on performance indicators of IQA algorithms", vol. 10, págs. 7507-7513, ene. de 2015.

[55] Media Technology and Innovation, *Empowering content creators with AI tools and game engines*, Document tech-I, Media technology and innovation, issue 51, Accessed: Nov. 17, 2023, 2022.

[56] TM Broadcast, *RTVE experimenta con Vídeo-MOS la automatización de la evaluación de la experiencia del espectador*, TM Broadcast, Accessed: Nov. 17, 2023, 2021.

[57] J. You, U. Reiter, M. M. Hannuksela, M. Gabbouj y A. Perkis, "Perceptual-based quality assessment for audio–visual services: A survey", *Signal Processing: Image Communication*, vol. 25, n.º 7, págs. 482-501, 2010, Special Issue on Image and Video Quality Assessment. DOI: https://doi.org/10.1016/j.image.2010.02.002.

[58] S. Winkler y C. Faller, "Perceived Audiovisual Quality of Low-Bitrate Multimedia Content", *IEEE Transactions on Multimedia*, vol. 8, n.º 5, págs. 973-980, 2006. DOI: 10.1109/TMM.2006.879871.

[59] R. Steinmetz, "Human perception of jitter and media synchronization", *IEEE Journal on Selected Areas in Communications*, vol. 14, n.º 1, págs. 61-72, 1996. DOI: 10.1109/49.481694.

[60] International Telecommunication Union, *SERIES E: OVERALL NETWORK OPERATION, TELEPHONE SERVICE, SERVICE OPERATION AND HUMAN FACTORS. Quality of telecommunication services: concepts, models, objectives and dependability planning – Terms and definitions related to the quality of telecommunication services. Definitions of terms related to quality of service: Recommendation ITU-T E.800 (09/2008)*, ITU Recommendation, Accessed: Nov. 17, 2023, 2008.

[61] ETSI, *Measurement guidelines for DVB systems - ETSI TR 101 290 V1.4.1*, ETSI Technical Report, Accessed: Apr. 9, 2025, 2020.

[62] K. Brunnström, K. De Moor, A. Dooms et al., *Qualinet White Paper on Definitions of Quality of Experience*. mar. de 2013.

[63] K. Zeng, T. Zhao, A. Rehman y Z. Wang, "Characterizing perceptual artifacts in compressed video streams", en *Human Vision and Electronic Imaging XIX*, B. E. Rogowitz, T. N. Pappas y H. de Ridder, eds., International Society for Optics y Photonics, vol. 9014, SPIE, 2014, 90140Q. DOI: 10.1117/12.2043128.

[64] A. Wahab, N. Ahmad y J. Schormans, "Variation in QoE of Passive Gaming Video Streaming for Different Packet Loss Ratios", en *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, 2020, págs. 1-4. DOI: 10.1109/QoMEX48832.2020.9123071.

[65] T. Hos¨feld, R. Schatz, T. Zinner, M. Seufert y P. Tran-Gia, "Transport protocol influences on YouTube videostreaming QoE", Fakultät für Angewandte Informatik, inf. téc., 2011, pág. 34.

[66] Š. Mrvelj y M. Matulin, "Impact of packet loss on the perceived quality of UDP-based multimedia streaming: a study of user quality of experience in real-life environments", *Multimedia Systems*, vol. 24, págs. 33-53, feb. de 2018. DOI: 10.1007/s00530-016-0531-8.

[67] A. C. Dalal, A. K. Bouchard, S. Cantor, Y. Guo y A. Johnson, "Assessing QoE of on-demand TCP video streams in real time", en *2012 IEEE International Conference on Communications (ICC)*, 2012, págs. 1165-1170. DOI: 10.1109/ICC.2012.6364073.

[68] K. D. Singh, Y. Hadjadj-Aoul y G. Rubino, "Quality of experience estimation for adaptive HTTP/TCP video streaming using H.264/AVC", en *2012 IEEE Consumer Communications and Networking Conference (CCNC)*, 2012, págs. 127-131. DOI: 10.1109/CCNC.2012.6181070.

[69] N. Barman y M. G. Martini, "QoE Modeling for HTTP Adaptive Video Streaming–A Survey and Open Challenges", *IEEE Access*, vol. 7, págs. 30 831-30 859, 2019. DOI: 10.1109/ACCESS.2019.2901778.

[70] H. G. Msakni y H. Youssef, "Ensuring video QoE using HTTP Adaptive Streaming: Issues and challenges", en *2016 5th International Conference on Multimedia Computing and Systems (ICMCS)*, 2016, págs. 200-205. DOI: 10.1109/ICMCS.2016.7905586.

[71] International Telecommunication Union, *Recommendation ITU-T H.264: Advanced video coding for generic audiovisual services*, 2015.

[72] International Telecommunication Union, *Recommendation ITU-T H.265: High efficiency video coding*, 2013.

[73] H. Schwarz, M. Coban, M. Karczewicz et al., "Quantization and Entropy Coding in the Versatile Video Coding (VVC) Standard", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, n.º 10, págs. 3891-3906, 2021. DOI: 10.1109/TCSVT.2021.3072202.

[74] X. Jin e Y. Chai, "Research on Quantization Parameter Decision Scheme for High Efficiency Video Coding", *Applied Sciences*, vol. 13, n.º 23, 2023. DOI: 10.3390/app132312758.

[75] D. Petreski y T. Kartalov, "Next Generation Video Compression Standards – Performance Overview", en *2023 30th International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2023, págs. 1-5. DOI: 10.1109/IWSSIP58668.2023.10180261.

[76] W. Lin, L. Dong y P. Xue, "Visual distortion gauge based on discrimination of noticeable contrast changes", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, n.º 7, págs. 900-909, 2005. DOI: 10.1109/TCSVT.2005.848345.

[77] H. Wu y M. Yuen, "A generalized block-edge impairment metric for video coding", *IEEE Signal Processing Letters*, vol. 4, n.º 11, págs. 317-320, 1997. DOI: 10.1109/97.641398.

[78] R. Pastrana-Vidal, J. Gicquel, C. Colomes y H. Cherifi, "Sporadic frame dropping impact on quality perception", *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 5292, jun. de 2004. DOI: 10.1117/12.525746.

[79] U. Reiter, K. Brunnström, K. De Moor et al., "Factors Influencing Quality of Experience", en *Quality of Experience: Advanced Concepts, Applications and Methods*, S. Möller y A. Raake, eds. Cham: Springer International Publishing, 2014, págs. 55-72. DOI: 10.1007/978-3-319-02681-7_4.

[80] T. Konaszyński, D. Juszka y M. Leszczuk, "Impact of the Stimulus Presentation Structure on Subjective Video Quality Assessment", *Electronics*, vol. 12, n.º 22, 2023. DOI: 10.3390/electronics12224593.

[81] F. Z. Allan, H. Bousbia-Salah y L. Hamami, "State of the art on human visual system modeling",

[82] P. Reichl, S. Egger, R. Schatz y A. D'Alconzo, "The Logarithmic Nature of QoE and the Role of the Weber-Fechner Law in QoE Assessment", en *2010 IEEE International Conference on Communications*, 2010, págs. 1-5. DOI: 10.1109/ICC.2010.5501894.

[83] A. Harijan, "Effect of Luminance Adaptation on Contrast Threshold in HVS", Tesis de mtría., Itä-Suomen yliopisto, 2023.

[84] M. Bhat, J.-M. Thiesse y P. L. Callet, "HVS based perceptual pre-processing for video coding", en *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019, págs. 1-5. DOI: 10.23919/EUSIPCO.2019.8903172.

[85] D. Yuan, T. Zhao, Y. Xu, H. Xue y L. Lin, "Visual JND: A Perceptual Measurement in Video Coding", *IEEE Access*, vol. 7, págs. 29 014-29 022, 2019. DOI: 10.1109/ACCESS.2019.2901342.

[86] H. Wang, X. Zhang, C. Yang y C.-C. J. Kuo, "Analysis and Prediction of JND-Based Video Quality Model", en *2018 Picture Coding Symposium (PCS)*, 2018, págs. 278-282. DOI: 10.1109/PCS.2018.8456243.

[87] H. Wang, I. Katsavounidis, X. Zhang, C. Yang y C.-C. J. Kuo, "A user model for JND-based video quality assessment: theory and applications", en *Applications of Digital Image Processing XLI*, A. G. Tescher, ed., International Society for Optics y Photonics, vol. 10752, SPIE, 2018, pág. 107520M. DOI: 10.1117/12.2320813.

[88] H. Amirpour, J. Zhu, R. Schatz, P. Le Callet y C. Timmerer, "Exploring Bitrate Costs for Enhanced User Satisfaction: A Just Noticeable Difference (JND) Perspective", en *2024 Data Compression Conference (DCC)*, 2024, págs. 432-441. DOI: 10.1109/DCC58796.2024.00051.

[89] J. You, A. Perkis y M. Gabbouj, "Improving image quality assessment with modeling visual attention", en *2010 2nd European Workshop on Visual Information Processing (EUVIP)*, 2010, págs. 177-182. DOI: 10.1109/EUVIP.2010.5699102.

[90] L. Itti, C. Koch y E. Niebur, "A model of saliency-based visual attention for rapid scene analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, n.º 11, págs. 1254-1259, 1998. DOI: 10.1109/34.730558.

[91] C. Li, J. Xue, N. Zheng, X. Lan y Z. Tian, "Spatio-Temporal Saliency Perception via Hypercomplex Frequency Spectral Contrast", *Sensors*, vol. 13, n.º 3, págs. 3409-3431, 2013. DOI: 10.3390/s130303409.

[92] A. Borji, M.-M. Cheng, Q. Hou, H. Jiang y J. Li, "Salient object detection: A survey", *Computational Visual Media*, vol. 5, n.º 2, págs. 117-150, 2019. DOI: 10.1007/s41095-019-0149-9.

[93] A. Imran, F. Guraya y F. Alaya Cheikh, "A visual attention based reference free perceptual quality metric", ago. de 2010, págs. 55-60. DOI: 10.1109/EUVIP.2010.5699132.

[94] L. Wei, N. Sang e Y. Wang, "A spatiotemporal saliency model of visual attention based on maximum entropy", *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, vol. 38, ene. de 2010.

[95] A. Rahman, G. Song, D. Pellerin y D. Houzet, "Spatio-temporal fusion of visual attention model", en *2011 19th European Signal Processing Conference*, 2011, págs. 2029-2033.

[96] W. Kim, C. Jung y C. Kim, "Spatiotemporal Saliency Detection and Its Applications in Static and Dynamic Scenes", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, n.º 4, págs. 446-456, 2011. DOI: 10.1109/TCSVT.2011.2125450.

[97] A. Rahman, D. Houzet, D. Pellerin y L. Agud, "GPU implementation of motion estimation for visual saliency", en *2010 Conference on Design and Architectures for Signal and Image Processing (DASIP)*, 2010, págs. 222-227. DOI: 10.1109/DASIP.2010.5706268.

[98] Y. Zhai y M. Shah, "Visual attention detection in video sequences using spatiotemporal cues", oct. de 2006, págs. 815-824. DOI: 10.1145/1180639.1180824.

[99] M. A. Hoque, T. Islam, T. Ahmed y A. Amin, "Autonomous Face Detection System from Real-time Video Streaming for Ensuring the Intelligence Security System", en *2020*

*6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2020, págs. 261-265. DOI: 10.1109/ICACCS48705.2020.9074260.

[100] A. Rahman, D. Houzet, D. Pellerin, S. Marat y N. Guyader, "Parallel implementation of a spatio-temporal visual saliency model", *Journal of Real-Time Image Processing*, vol. 6, n.º 1, págs. 3-14, 2011. DOI: 10.1007/s11554-010-0164-7.

[101] A. R. Patrone, C. Valuch, U. Ansorge y O. Scherzer, *Dynamical optical flow of saliency maps for predicting visual attention*, 2016. arXiv: 1606.07324 [cs.CV].

[102] J. Kuang, G. M. Johnson y M. D. Fairchild, "Image Preference Scaling for HDR Image Rendering", *Color and Imaging Conference*, vol. 13, n.º 1, págs. 8-8, 2005. DOI: 10.2352/CIC.2005.13.1.art00002.

[103] J. Jang, H. Jang, T. Eo, K. Bang y D. Hwang, "No-reference Automatic Quality Assessment for Colorfulness-Adjusted, Contrast-Adjusted, and Sharpness-Adjusted Images Using High-Dynamic-Range-Derived Features", *Applied Sciences*, vol. 8, n.º 9, 2018. DOI: 10.3390/app8091688.

[104] C. Li y T. Chen, "Aesthetic Visual Quality Assessment of Paintings", *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, n.º 2, págs. 236-252, 2009. DOI: 10.1109/JSTSP.2009.2015077.

[105] R. Datta, J. Li y J. Z. Wang, "Algorithmic inferencing of aesthetics and emotion in natural images: An exposition", en *2008 15th IEEE International Conference on Image Processing*, 2008, págs. 105-108. DOI: 10.1109/ICIP.2008.4711702.

[106] A. K. Moorthy, P. Obrador y N. Oliver, "Towards Computational Models of the Visual Aesthetic Appeal of Consumer Videos", en *Computer Vision – ECCV 2010*, K. Daniilidis, P. Maragos y N. Paragios, eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, págs. 1-14.

[107] ITU-R, *Methodologies for the subjective assessment of the quality of television images, document Rec. BT.500*, ITU-R Document, Accessed: Nov. 17, 2023, 2023.

[108] ITU-T, *Subjective video quality assessment methods for multimedia applications, document Rec. P.910*, ITU-T Document, Accessed: Nov. 17, 2023, 2022.

[109] H. Sheikh, M. Sabir y A. Bovik, "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms", *IEEE Transactions on Image Processing*, vol. 15, n.º 11, págs. 3440-3451, 2006. DOI: 10.1109/TIP.2006.881959.

[110] A. Ninassi, P. L. Callet y F. Autrusseau, "Pseudo no reference image quality metric using perceptual data hiding", en *Human Vision and Electronic Imaging XI*, B. E. Rogowitz, T. N. Pappas y S. J. Daly, eds., International Society for Optics y Photonics, vol. 6057, SPIE, 2006, 60570G. DOI: 10.1117/12.650780.

[111] E. C. Larson y D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy", *Journal of Electronic Imaging*, vol. 19, n.º 1, pág. 011 006, 2010. DOI: 10.1117/1.3267105.

[112] H. Lin, V. Hosu y D. Saupe, "KADID-10k: A Large-scale Artificially Distorted IQA Database", en *2019 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, IEEE, 2019, págs. 1-3.

[113] D. Kundu, D. Ghadiyaram, A. C. Bovik y B. L. Evans, "Large-Scale Crowdsourced Study for Tone-Mapped HDR Pictures", *IEEE Transactions on Image Processing*, vol. 26, n.º 10, págs. 4725-4740, 2017. DOI: 10.1109/TIP.2017.2713945.

[114] K. Seshadrinathan, R. Soundararajan, A. C. Bovik y L. K. Cormack, "Study of Subjective and Objective Quality Assessment of Video", *IEEE Transactions on Image Processing*, vol. 19, n.º 6, págs. 1427-1441, 2010. DOI: 10.1109/TIP.2010.2042111.

[115] A. K. Moorthy, L. K. Choi, A. C. Bovik y G. de Veciana, "Video Quality Assessment on Mobile Devices: Subjective, Behavioral and Objective Studies", *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, n.º 6, págs. 652-671, 2012. DOI: 10.1109/JSTSP.2012.2212417.

[116] S. Péchard, R. Pépion y P. Le Callet, "Suitable methodology in subjective video quality assessment: a resolution dependent paradigm", en *International Workshop on Image Media Quality and its Applications, IMQA2008*, Kyoto, Japan, sep. de 2008, pág. 6.

[117] Y. Wang, S. Inguva y B. Adsumilli, "YouTube UGC Dataset for Video Compression Research", en *2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP)*, IEEE, sep. de 2019. DOI: 10.1109/mmsp.2019.8901772.

[118] A. Mercat, M. Viitanen y J. Vanne, "UVG dataset: 50/120fps 4K sequences for video codec analysis and development", en *Proceedings of the 11th ACM Multimedia Systems Conference*, ép. MMSys '20, Istanbul, Turkey: Association for Computing Machinery, 2020, págs. 297-302. DOI: 10.1145/3339825.3394937.

[119] L. Song, X. Tang, W. Zhang, X. Yang y P. Xia, "The SJTU 4K video sequence dataset", jul. de 2013, págs. 34-35. DOI: 10.1109/QoMEX.2013.6603201.

[120] V. Hosu, F. Hahn, M. Jenadeleh et al., "The Konstanz natural video database (KoNViD-1k)", en *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, 2017, págs. 1-6. DOI: 10.1109/QoMEX.2017.7965673.

[121] F. Zhang, F. Mercer Moss, R. Baddeley y D. Bull, "BVI-HD: A Video Quality Database for HEVC Compressed and Texture Synthesised Content", *IEEE Transactions on Multimedia*, vol. PP, págs. 1-1, mar. de 2018. DOI: 10.1109/TMM.2018.2817070.

[122] M. H. Pinson, "ITS4S: A Video Quality Dataset with Four-Second Unrepeated Scenes", 2018.

[123] C. Keimel, J. Habigt, T. Habigt, M. Rothbucher y K. Diepold, "Visual quality of current coding technologies at high definition IPTV bitrates", en *2010 IEEE International Workshop on Multimedia Signal Processing*, 2010, págs. 390-393. DOI: 10.1109/MMSP.2010.5662052.

[124] W. Lin y C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey", *Journal of Visual Communication and Image Representation*, vol. 22, n.º 4, págs. 297-312, 2011. DOI: https://doi.org/10.1016/j.jvcir.2011.01.005.

[125] M. Vranješ, S. Rimac-Drlje y K. Grgić, "Review of objective video quality metrics and performance comparison using different databases", *Signal Processing: Image Communication*, vol. 28, n.º 1, págs. 1-19, 2013. DOI: https://doi.org/10.1016/j.image.2012.10.003.

[126] M. Vranjes, S. Rimac-Drlje y D. Zagar, "Objective video quality metrics", en *ELMAR 2007*, 2007, págs. 45-49. DOI: 10.1109/ELMAR.2007.4418797.

[127] Z. Wang, A. Bovik, H. Sheikh y E. Simoncelli, "Image quality assessment: from error visibility to structural similarity", *IEEE Transactions on Image Processing*, vol. 13, n.º 4, págs. 600-612, 2004. DOI: 10.1109/TIP.2003.819861.

[128] C. G. Bampis, Z. Li y A. C. Bovik, "Spatiotemporal Feature Integration and Model Fusion for Full Reference Video Quality Assessment", *IEEE Transactions on Circuits*

*and Systems for Video Technology*, vol. 29, n.º 8, págs. 2256-2270, 2019. DOI: 10.1109/TCSVT.2018.2868262.

[129] K. Seshadrinathan y A. C. Bovik, "Motion Tuned Spatio-Temporal Quality Assessment of Natural Videos", *IEEE Transactions on Image Processing*, vol. 19, n.º 2, págs. 335-350, 2010. DOI: 10.1109/TIP.2009.2034992.

[130] R. Soundararajan y A. C. Bovik, "Video Quality Assessment by Reduced Reference Spatio-Temporal Entropic Differencing", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, n.º 4, págs. 684-694, 2013. DOI: 10.1109/TCSVT.2012.2214933.

[131] C. Feng, D. Danier, F. Zhang y D. Bull, *RankDVQA: Deep VQA based on Ranking-inspired Hybrid Training*, 2023. arXiv: 2202.08595 [eess.IV].

[132] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola y V. Lukin, "On between-coefficient contrast masking of DCT basis functions", *Proc of the 3rd Int Workshop on Video Processing and Quality Metrics for Consumer Electronics*, ene. de 2007.

[133] Z. Wang, E. Simoncelli y A. Bovik, "Multiscale structural similarity for image quality assessment", en *The Thrity-Seventh Asilomar Conference on Signals, Systems and Computers, 2003*, vol. 2, 2003, 1398-1402 Vol.2. DOI: 10.1109/ACSSC.2003.1292216.

[134] D. M. Chandler y S. S. Hemami, "VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images", *IEEE Transactions on Image Processing*, vol. 16, n.º 9, págs. 2284-2298, 2007. DOI: 10.1109/TIP.2007.901820.

[135] A. C. Bovik, "A VISUAL INFORMATION FIDELITY APPROACH TO VIDEO QUALITY ASSESSMENT", 2005.

[136] L. Zhang, L. Zhang, X. Mou y D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assessment", *IEEE Transactions on Image Processing*, vol. 20, n.º 8, págs. 2378-2386, 2011. DOI: 10.1109/TIP.2011.2109730.

[137] S. Rimac-Drlje, M. Vranješ y D. Žagar, "Foveated mean squared error—a novel video quality metric", *Multimedia Tools and Applications*, vol. 49, págs. 425-445, 2010. DOI: 10.1007/s11042-009-0442-1.

[138] P. V. Vu y D. M. Chandler, "ViS3: an algorithm for video quality assessment via analysis of spatial and spatiotemporal slices", *Journal of Electronic Imaging*, vol. 23, n.º 1, pág. 013 016, 2014. DOI: 10.1117/1.JEI.23.1.013016.

[139] P. V. Vu, C. T. Vu y D. M. Chandler, "A spatiotemporal most-apparent-distortion model for video quality assessment", en *2011 18th IEEE International Conference on Image Processing*, 2011, págs. 2505-2508. DOI: 10.1109/ICIP.2011.6116171.

[140] F. Zhang y D. R. Bull, "A Perception-Based Hybrid Model for Video Quality Assessment", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, n.º 6, págs. 1017-1028, 2016. DOI: 10.1109/TCSVT.2015.2428551.

[141] K. Manasa y S. S. Channappayya, "An Optical Flow-Based Full Reference Video Quality Assessment Algorithm", *IEEE Transactions on Image Processing*, vol. 25, n.º 6, págs. 2480-2492, 2016. DOI: 10.1109/TIP.2016.2548247.

[142] J. Wu, Y. Liu, W. Dong, G. Shi y W. Lin, "Quality Assessment for Video With Degradation Along Salient Trajectories", *IEEE Transactions on Multimedia*, vol. 21, n.º 11, págs. 2738-2749, 2019. DOI: 10.1109/TMM.2019.2908377.

[143] P. C. Madhusudana, N. Birkbeck, Y. Wang, B. Adsumilli y A. C. Bovik, "ST-GREED: Space-Time Generalized Entropic Differences for Frame Rate Dependent Video Quality Prediction", *IEEE Transactions on Image Processing*, vol. 30, págs. 7446-7457, 2021. DOI: 10.1109/TIP.2021.3106801.

[144] P. G. Freitas, W. Y. Akamine y M. C. Farias, "Using multiple spatio-temporal features to estimate video quality", *Signal Processing: Image Communication*, vol. 64, págs. 1-10, 2018. DOI: https://doi.org/10.1016/j.image.2018.02.010.

[145] S. V. R. Dendi, G. Krishnappa y S. S. Channappayya, "Full-Reference Video Quality Assessment Using Deep 3D Convolutional Neural Networks", en *2019 National Conference on Communications (NCC)*, 2019, págs. 1-5. DOI: 10.1109/NCC.2019.8732265.

[146] W. Kim, J. Kim, S. Ahn, J. Kim y S. Lee, "Deep Video Quality Assessor: From Spatio-Temporal Visual Sensitivity to a Convolutional Neural Aggregation Network", en *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu e Y. Weiss, eds., Cham: Springer International Publishing, 2018, págs. 224-241.

[147] M. Xu, J. Chen, H. Wang, S. Liu, G. Li y Z. Bai, "C3DVQA: Full-Reference Video Quality Assessment with 3D Convolutional Neural Network", en *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, págs. 4447-4451. DOI: 10.1109/ICASSP40776.2020.9053031.

[148] J. Chen, H. Wang, M. Xu, G. Li y S. Liu, "Deep Neural Networks for End-to-End Spatiotemporal Video Quality Prediction and Aggregation", en *2021 IEEE International Conference on Multimedia and Expo (ICME)*, 2021, págs. 1-6. DOI: 10.1109/ICME51207.2021.9428209.

[149] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand y W. Samek, "Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment", *IEEE Transactions on Image Processing*, vol. 27, n.º 1, págs. 206-219, 2018. DOI: 10.1109/TIP.2017.2760518.

[150] K. Ding, K. Ma, S. Wang y E. P. Simoncelli, "Image Quality Assessment: Unifying Structure and Texture Similarity", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, n.º 5, págs. 2567-2581, 2022. DOI: 10.1109/TPAMI.2020.3045810.

[151] J. Kim y S. Lee, "Deep Learning of Human Visual Sensitivity in Image Quality Assessment Framework", en *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, págs. 1969-1977. DOI: 10.1109/CVPR.2017.213.

[152] P. C. Madhusudana, N. Birkbeck, Y. Wang, B. Adsumilli y A. C. Bovik, "Image Quality Assessment Using Contrastive Learning", *IEEE Transactions on Image Processing*, vol. 31, págs. 4149-4161, 2022. DOI: 10.1109/TIP.2022.3181496.

[153] C. G. Bampis, P. Gupta, R. Soundararajan y A. C. Bovik, "SpEED-QA: Spatial Efficient Entropic Differencing for Image and Video Quality", *IEEE Signal Processing Letters*, vol. 24, n.º 9, págs. 1333-1337, 2017. DOI: 10.1109/LSP.2017.2726542.

[154] Z. Parvez Sazzad, Y. Kawayoke e Y. Horita, "No reference image quality assessment for JPEG2000 based on spatial features", *Signal Processing: Image Communication*, vol. 23, n.º 4, págs. 257-268, 2008. DOI: https://doi.org/10.1016/j.image.2008.03.005.

[155] Z. Wang, A. Bovik y B. Evan, "Blind measurement of blocking artifacts in images", en *Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101)*, vol. 3, 2000, 981-984 vol.3. DOI: 10.1109/ICIP.2000.899622.

[156] A. Bovik y S. Liu, "DCT-domain blind measurement of blocking artifacts in DCT-coded images", en *2001 IEEE International Conference on Acoustics, Speech, and*

*Signal Processing. Proceedings (Cat. No.01CH37221)*, vol. 3, 2001, 1725-1728 vol.3. DOI: 10.1109/ICASSP.2001.941272.

[157] K. Tan y M. Ghanbari, "Blockiness detection for MPEG2-coded video", *IEEE Signal Processing Letters*, vol. 7, n.º 8, págs. 213-215, 2000. DOI: 10.1109/97.855443.

[158] T. Vlachos, "Detection of blocking artifacts in compressed video", *Electronics Letters*, vol. 36, págs. 1106-1108, jul. de 2000. DOI: 10.1049/el:20000847.

[159] S. Suthaharan, "Perceptual quality metric for digital video coding", *Electronics Letters*, vol. 39, págs. 431-433, abr. de 2003. DOI: 10.1049/el:20030308.

[160] J. Lu, "Image analysis for video artifact estimation and measurement", en *Machine Vision Applications in Industrial Inspection IX*, M. A. Hunt, ed., International Society for Optics y Photonics, vol. 4301, SPIE, 2001, págs. 166-174. DOI: 10.1117/12.420909.

[161] X. Zhu y P. Milanfar, "A no-reference sharpness metric sensitive to blur and noise", en *2009 International Workshop on Quality of Multimedia Experience*, 2009, págs. 64-69. DOI: 10.1109/QOMEX.2009.5246976.

[162] C. Chen, M. Izadi y A. Kokaram, "A Perceptual Quality Metric for Videos Distorted by Spatially Correlated Noise", en *Proceedings of the 24th ACM International Conference on Multimedia*, ép. MM '16, Amsterdam, The Netherlands: Association for Computing Machinery, 2016, págs. 1277-1285. DOI: 10.1145/2964284.2964302.

[163] A. Norkin y N. Birkbeck, "Film Grain Synthesis for AV1 Video Codec", en *2018 Data Compression Conference*, 2018, págs. 3-12. DOI: 10.1109/DCC.2018.00008.

[164] H. Liu, N. Klomp e I. Heynderickx, "A No-Reference Metric for Perceived Ringing Artifacts in Images", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, n.º 4, págs. 529-539, 2010. DOI: 10.1109/TCSVT.2009.2035848.

[165] X. Feng y J. P. Allebach, "Measurement of ringing artifacts in JPEG images", en *Digital Publishing*, J. P. Allebach y H. Chao, eds., International Society for Optics y Photonics, vol. 6076, SPIE, 2006, 60760A. DOI: 10.1117/12.645089.

[166] Z. Tu, J. Lin, Y. Wang, B. Adsumilli y A. C. Bovik, "BBAND INDEX: A NO-REFERENCE BANDING ARTIFACT PREDICTOR", en *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, págs. 2712-2716. DOI: 10.1109/ICASSP40776.2020.9053634.

[167] S. Saini, A. Saha y A. C. Bovik, *HIDRO-VQA: High Dynamic Range Oracle for Video Quality Assessment*, 2023. arXiv: 2311.11059 [cs.CV].

[168] K.-C. Yang, C. C. Guest, K. El-Maleh y P. K. Das, "Perceptual Temporal Quality Metric for Compressed Video", *IEEE Transactions on Multimedia*, vol. 9, n.º 7, págs. 1528-1535, 2007. DOI: 10.1109/TMM.2007.906576.

[169] R. Pastrana-Vidal, J.-C. Gicquel y F. Telecom, "Automatic quality assessment of video fluidity impairments using a No-Reference Metric", ene. de 2006.

[170] Q. Huynh-Thu y M. Ghanbari, "Impact of jitter and jerkiness on perceived video quality", ene. de 2006.

[171] Y.-F. Ou, Z. Ma, T. Liu e Y. Wang, "Perceptual Quality Assessment of Video Considering Both Frame Rate and Quantization Artifacts", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, n.º 3, págs. 286-298, 2011. DOI: 10.1109/TCSVT.2010.2087833.

[172] E. Ong, S. Wu, M. Loke et al., "Video quality monitoring of streamed videos", abr. de 2009, págs. 1153-1156. DOI: 10.1109/ICASSP.2009.4959793.

[173] M. A. Saad y A. C. Bovik, "Blind quality assessment of videos using a model of natural scene statistics and motion coherency", en *2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2012, págs. 332-336. DOI: 10.1109/ACSSC.2012.6489018.

[174] J. Caviedes y F. Oberti, "No-Reference Quality Metric for Degraded and Enhanced Video", vol. 5150, ene. de 2003, págs. 621-632. DOI: 10.1201/9781420027822.ch10.

[175] M. Farias y S. Mitra, "No-reference video quality metric based on artifact measurements", en *IEEE International Conference on Image Processing 2005*, vol. 3, 2005, págs. III-141. DOI: 10.1109/ICIP.2005.1530348.

[176] F. Massidda, D. D. Giusto y C. Perra, "No reference video quality estimation based on human visual system for 2.5/3G devices", en *Human Vision and Electronic Imaging X*, B. E. Rogowitz, T. N. Pappas y S. J. Daly, eds., International Society for Optics y Photonics, vol. 5666, SPIE, 2005, págs. 168-179. DOI: 10.1117/12.594032.

[177] R. Babu, B. Ajit, S. Bopardikar, A. Perkis e I. Hillestad, "No-Reference metrics for video streaming applications", *International Workshop on Packet Video*, ene. de 2004.

[178] R. Dosselmann y X. Dong Yang, "A Prototype No-Reference Video Quality System", en *Fourth Canadian Conference on Computer and Robot Vision (CRV '07)*, 2007, págs. 411-417. DOI: 10.1109/CRV.2007.6.

[179] M. H. Pinson, "Why No Reference Metrics for Image and Video Quality Lack Accuracy and Reproducibility", *IEEE Transactions on Broadcasting*, vol. 69, n.º 1, págs. 97-117, 2023. DOI: 10.1109/TBC.2022.3191059.

[180] D. L. Ruderman, "The statistics of natural images", *Network: Computation in Neural Systems*, vol. 5, n.º 4, págs. 517-548, 1994. DOI: 10.1088/0954-898X_5_4_006. eprint: https://doi.org/10.1088/0954-898X_5_4_006.

[181] A. Mittal, R. Soundararajan y A. C. Bovik, "Making a "Completely Blind" Image Quality Analyzer", *IEEE Signal Processing Letters*, vol. 20, n.º 3, págs. 209-212, 2013. DOI: 10.1109/LSP.2012.2227726.

[182] A. Mittal, A. K. Moorthy y A. C. Bovik, "No-Reference Image Quality Assessment in the Spatial Domain", *IEEE Transactions on Image Processing*, vol. 21, n.º 12, págs. 4695-4708, 2012. DOI: 10.1109/TIP.2012.2214050.

[183] W. Xue, X. Mou, L. Zhang, A. C. Bovik y X. Feng, "Blind Image Quality Assessment Using Joint Statistics of Gradient Magnitude and Laplacian Features", *IEEE Transactions on Image Processing*, vol. 23, n.º 11, págs. 4850-4862, 2014. DOI: 10.1109/TIP.2014.2355716.

[184] Y. Zhang y D. M. Chandler, "No-reference image quality assessment based on log-derivative statistics of natural scenes", *Journal of Electronic Imaging*, vol. 22, n.º 4, pág. 043 025, 2013. DOI: 10.1117/1.JEI.22.4.043025.

[185] D. Kundu, D. Ghadiyaram, A. C. Bovik y B. L. Evans, "No-reference image quality assessment for high dynamic range images", en *2016 50th Asilomar Conference on Signals, Systems and Computers*, 2016, págs. 1847-1852. DOI: 10.1109/ACSSC.2016.7869704.

[186] M. A. Saad, A. C. Bovik y C. Charrier, "A DCT Statistics-Based Blind Image Quality Index", *IEEE Signal Processing Letters*, vol. 17, n.º 6, págs. 583-586, 2010. DOI: 10.1109/LSP.2010.2045550.

[187] M. A. Saad, A. C. Bovik y C. Charrier, "Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain", *IEEE Transactions on Image Processing*, vol. 21, n.º 8, págs. 3339-3352, 2012. DOI: 10.1109/TIP.2012.2191563.

[188] A. K. Moorthy y A. C. Bovik, "A Two-Step Framework for Constructing Blind Image Quality Indices", *IEEE Signal Processing Letters*, vol. 17, n.º 5, págs. 513-516, 2010. DOI: 10.1109/LSP.2010.2043888.

[189] A. K. Moorthy y A. C. Bovik, "Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality", *IEEE Transactions on Image Processing*, vol. 20, n.º 12, págs. 3350-3364, 2011. DOI: 10.1109/TIP.2011.2147325.

[190] D. Ghadiyaram y A. C. Bovik, "Perceptual quality prediction on authentically distorted images using a bag of features approach", *Journal of Vision*, vol. 17, n.º 1, págs. 32-32, ene. de 2017. DOI: 10.1167/17.1.32. eprint: https://arvojournals.org/arvo/content_public/journal/jov/935953/i1534-7362-17-1-32.pdf.

[191] P. Ye, J. Kumar, L. Kang y D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment", en *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, págs. 1098-1105. DOI: 10.1109/CVPR.2012.6247789.

[192] Z. Tu, Y. Wang, N. Birkbeck, B. Adsumilli y A. C. Bovik, "UGC-VQA: Benchmarking Blind Video Quality Assessment for User Generated Content", *IEEE Transactions on Image Processing*, vol. 30, págs. 4449-4464, 2021. DOI: 10.1109/TIP.2021.3072221.

[193] M. A. Saad, A. C. Bovik y C. Charrier, "Blind Prediction of Natural Video Quality", *IEEE Transactions on Image Processing*, vol. 23, n.º 3, págs. 1352-1365, 2014. DOI: 10.1109/TIP.2014.2299154.

[194] A. Mittal, M. A. Saad y A. C. Bovik, "A Completely Blind Video Integrity Oracle", *IEEE Transactions on Image Processing*, vol. 25, n.º 1, págs. 289-300, 2016. DOI: 10.1109/TIP.2015.2502725.

[195] X. Li, Q. Guo y X. Lu, "Spatiotemporal Statistics for Video Quality Assessment", *IEEE Transactions on Image Processing*, vol. 25, n.º 7, págs. 3329-3342, 2016. DOI: 10.1109/TIP.2016.2568752.

[196] H. Men, H. Lin y D. Saupe, "Spatiotemporal Feature Combination Model for No-Reference Video Quality Assessment", en *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, 2018, págs. 1-3. DOI: 10.1109/QoMEX.2018.8463426.

[197] J. P. Ebenezer, Z. Shang, Y. Wu, H. Wei y A. C. Bovik, "No-Reference Video Quality Assessment Using Space-Time Chips", en *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*, 2020, págs. 1-6. DOI: 10.1109/MMSP48831.2020.9287151.

[198] J. Korhonen, "Two-Level Approach for No-Reference Consumer Video Quality Assessment", *IEEE Transactions on Image Processing*, vol. 28, n.º 12, págs. 5923-5938, 2019. DOI: 10.1109/TIP.2019.2923051.

[199] Z. Ying, M. Mandal, D. Ghadiyaram y A. Bovik, "Patch-VQ: 'Patching Up' the Video Quality Problem", en *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, págs. 14 014-14 024. DOI: 10.1109/CVPR46437.2021.01380.

[200] D. Varga, "No-Reference Video Quality Assessment Using Multi-Pooled, Saliency Weighted Deep Features and Decision Fusion", *Sensors*, vol. 22, n.º 6, 2022. DOI: 10.3390/s22062209.

[201] B. Chen, L. Zhu, G. Li, F. Lu, H. Fan y S. Wang, "Learning Generalized Spatial-Temporal Deep Feature Representation for No-Reference Video Quality Assessment", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, n.º 4, págs. 1903-1916, 2022. DOI: 10.1109/TCSVT.2021.3088505.

[202] W. Zhou y Z. Chen, "Deep Local and Global Spatiotemporal Feature Aggregation for Blind Video Quality Assessment", en *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, 2020, págs. 338-341. DOI: 10.1109/VCIP49819.2020.9301764.

[203] J. Korhonen, Y. Su y J. You, "Blind Natural Video Quality Prediction via Statistical Temporal Features and Deep Spatial Features", en *Proceedings of the 28th ACM International Conference on Multimedia*, ép. MM '20, Seattle, WA, USA: Association for Computing Machinery, 2020, págs. 3311-3319. DOI: 10.1145/3394171.3413845.

[204] Z. Tu, X. Yu, Y. Wang, N. Birkbeck, B. Adsumilli y A. C. Bovik, "RAPIQUE: Rapid and Accurate Video Quality Prediction of User Generated Content", *IEEE Open Journal of Signal Processing*, vol. 2, págs. 425-440, 2021. DOI: 10.1109/OJSP.2021.3090333.

[205] D. Li, T. Jiang y M. Jiang, "Quality Assessment of In-the-Wild Videos", en *Proceedings of the 27th ACM International Conference on Multimedia*, ép. MM '19, Nice, France: Association for Computing Machinery, 2019, págs. 2351-2359. DOI: 10.1145/3343031.3351028.

[206] D. Liu, T. Jiang y M. Jiang, "Unified Quality Assessment of in-the-Wild Videos with Mixed Datasets Training", *International Journal of Computer Vision*, vol. 129, págs. 1238-1257, 2021. DOI: 10.1007/s11263-020-01408-w.

[207] W. Liu, Z. Duanmu y Z. Wang, "End-to-End Blind Quality Assessment of Compressed Videos Using Deep Neural Networks", en *Proceedings of the 26th ACM International Conference on Multimedia*, ép. MM '18, Seoul, Republic of Korea: Association for Computing Machinery, 2018, págs. 546-554. DOI: 10.1145/3240508.3240643.

[208] L. Lin, Z. Wang, J. He, W. Chen, Y. Xu y T. Zhao, "Deep Quality Assessment of Compressed Videos: A Subjective and Objective Study", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, n.º 6, págs. 2616-2626, 2023. DOI: 10.1109/TCSVT.2022.3227039.

[209] Y. Li, L.-M. Po, C.-H. Cheung et al., "No-Reference Video Quality Assessment With 3D Shearlet Transform and Convolutional Neural Networks", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, n.º 6, págs. 1044-1057, 2016. DOI: 10.1109/TCSVT.2015.2430711.

[210] H. Wu, C. Chen, L. Liao et al., "DisCoVQA: Temporal Distortion-Content Transformers for Video Quality Assessment", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, n.º 9, págs. 4840-4854, 2023. DOI: 10.1109/TCSVT.2023.3249741.

[211] Y. Wang, J. Ke, H. Talebi et al., "Rich features for perceptual quality assessment of UGC videos", en *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, págs. 13 430-13 439. DOI: 10.1109/CVPR46437.2021.01323.

[212] B. Li, W. Zhang, M. Tian, G. Zhai y X. Wang, "Blindly Assess Quality of In-the-Wild Videos via Quality-Aware Pre-Training and Motion Perception", *IEEE Transactions*

*on Circuits and Systems for Video Technology*, vol. 32, n.º 9, págs. 5944-5958, 2022. DOI: 10.1109/TCSVT.2022.3164467.

[213] M. Agarla, L. Celona y R. Schettini, "An Efficient Method for No-Reference Video Quality Assessment", *Journal of Imaging*, vol. 7, n.º 3, 2021. DOI: 10.3390/jimaging7030055.

[214] Y. Fang, Z. Li, J. Yan, X. Sui y H. Liu, "Study of Spatio-Temporal Modeling in Video Quality Assessment", *IEEE Transactions on Image Processing*, vol. 32, págs. 2693-2702, 2023. DOI: 10.1109/TIP.2023.3272480.

[215] A. Vishwakarma y K. Bhurchandi, "No-Reference Video Quality Assessment Using Local Structural and Quality-Aware Deep Features", *IEEE Transactions on Instrumentation and Measurement*, vol. PP, págs. 1-1, ene. de 2023. DOI: 10.1109/TIM.2023.3273654.

[216] J. Ke, T. Zhang, Y. Wang, P. Milanfar y F. Yang, *MRET: Multi-resolution Transformer for Video Quality Assessment*, 2023. arXiv: 2303.07489 [cs.CV].

[217] D. Varga, "No-Reference Video Quality Assessment Using the Temporal Statistics of Global and Local Image Features", *Sensors*, vol. 22, n.º 24, 2022. DOI: 10.3390/s22249696.

[218] K. Zhao, K. Yuan, M. Sun y X. Wen, *Zoom-VQA: Patches, Frames and Clips Integration for Video Quality Assessment*, 2023. arXiv: 2304.06440 [cs.CV].

[219] H. Wu, C. Chen, J. Hou et al., *FAST-VQA: Efficient End-to-end Video Quality Assessment with Fragment Sampling*, 2022. arXiv: 2207.02595 [cs.CV].

[220] H. Wu, L. Liao, C. Chaofeng et al., *Disentangling Aesthetic and Technical Effects for Video Quality Assessment of User Generated Content*, nov. de 2022. DOI: 10.48550/arXiv.2211.04894.

[221] European Broadcasting Union (EBU), "EBU – Recommendation R132: Signal Quality in HDTV Production and Broadcast Services", European Broadcasting Union, Geneva, inf. téc., 2011, Guidelines for technical, operational & creative staff on how to achieve and maintain sufficient technical quality along the production chain.

[222] T. Chai y R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?– Arguments against avoiding RMSE in the literature", *Geoscientific Model Development*, vol. 7, págs. 1247-1250, jun. de 2014. DOI: 10.5194/gmd-7-1247-2014.

[223] C. Willmott y K. Matsuura, "Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance", *Climate Research*, vol. 30, pág. 79, dic. de 2005. DOI: 10.3354/cr030079.

[224] Z. Wang y A. C. Bovik, "Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures", *IEEE Signal Processing Magazine*, vol. 26, n.º 1, págs. 98-117, 2009. DOI: 10.1109/MSP.2008.930649.

[225] B. W. Tatler, "The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions.", *Journal of vision*, vol. 7 14, págs. 4.1-17, 2007.

[226] M. Bindemann, "Scene and screen center bias early eye movements in scene viewing", *Vision Research*, vol. 50, n.º 23, págs. 2577-2587, 2010, Vision Research Reviews. DOI: https://doi.org/10.1016/j.visres.2010.08.016.

[227] A. Borji y L. Itti, "State-of-the-Art in Visual Attention Modeling", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, págs. 185-207, 2013.

[228] R. Kalboussi, M. Abdellaoui y A. Douik, "Overview on visual attention and computational saliency", en *2021 International Conference on Control, Automation and Diagnosis (ICCAD)*, 2021, págs. 1-6. DOI: 10.1109/ICCAD52417.2021.9638759.

[229] M. Cornia, L. Baraldi, G. Serra y R. Cucchiara, "Predicting Human Eye Fixations via an LSTM-Based Saliency Attentive Model", *IEEE Transactions on Image Processing*, vol. 27, n.º 10, págs. 5142-5154, 2018. DOI: 10.1109/TIP.2018.2851672.

[230] D.-P. Fan, M.-M. Cheng, J.-J. Liu, S.-H. Gao, Q. Hou y A. Borji, *Salient Objects in Clutter: Bringing Salient Object Detection to the Foreground*, 2018. arXiv: 1803.06091 [cs.CV].

[231] J. Vlaović, M. Vranješ, D. Grabić y D. Samardžija, "Comparison of Objective Video Quality Assessment Methods on Videos With Different Spatial Resolutions", en *2019 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2019, págs. 287-292. DOI: 10.1109/IWSSIP.2019.8787324.

[232] U. Sara, M. Akter y M. Uddin, "Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study", *Journal of Computer and Communications*, vol. 7, págs. 8-18, 2019. DOI: 10.4236/jcc.2019.73002.

[233] R. C. Gonzalez y R. E. Woods, *Digital Image Processing*, 4th. New York, NY: Pearson, 2018, pág. 1168.

[234] B. Zatt, M. Porto, J. Scharcanski y S. Bampi, "Gop structure adaptive to the video content for efficient H.264/AVC encoding", en *2010 IEEE International Conference on Image Processing*, 2010, págs. 3053-3056. DOI: 10.1109/ICIP.2010.5651700.

[235] L. Krulikovská, J. Polec y M. Martinovič, *Adaptive Group of Pictures Structure Based On the Positions of Video Cuts*, ver. 16315, ene. de 2018. DOI: 10.5281/zenodo.1086981.

[236] ETSI, *Digital Video Broadcasting (DVB); Specification for Service Information (SI) in Digital Video Broadcasting (DVB) Systems*, ETSI EN 300 468 V1.17.1, Accessed: Feb. 14, 2025, 2020.

[237] Á. Llorente, J. G. Pérez, J. A. Rodrigo, D. Jiménez y J. M. Menéndez, "Efficient Computational Cost Saving in Video Processing for QoE Estimation", *IEEE Access*, vol. 12, págs. 34 846-34 862, 2024. DOI: 10.1109/ACCESS.2024.3373193.

[238] Á. Llorente, A. del Rio, Y. C. Semerci, J. A. Kurano, D. Jimenez y J. M. Menéndez, "Assessment of cognitive games to improve the quality of life of Parkinson's and Alzheimer's patients", *DIGITAL HEALTH*, vol. 10, pág. 20 552 076 241 254 733, 2024. DOI: 10.1177/20552076241254733. eprint: https://doi.org/10.1177/20552076241254733.

[239] R. Martínez, Á. Llorente, A. del Rio, J. Serrano y D. Jimenez, "Performance Evaluation of YOLOv8-Based Bib Number Detection in Media Streaming Race", *IEEE Transactions on Broadcasting*, vol. 70, n.º 3, págs. 1126-1138, 2024. DOI: 10.1109/TBC.2024.3414656.

[240] P. Schummer, A. del Río, J. Serrano, D. Jiménez, G. Sánchez y Á. Llorente, "Machine Learning-Based Network Anomaly Detection: Design, Implementation, and Evaluation", *AI*, 2024.

# Apéndice A

# Influencia de grafismos en las métricas de evaluación de calidad

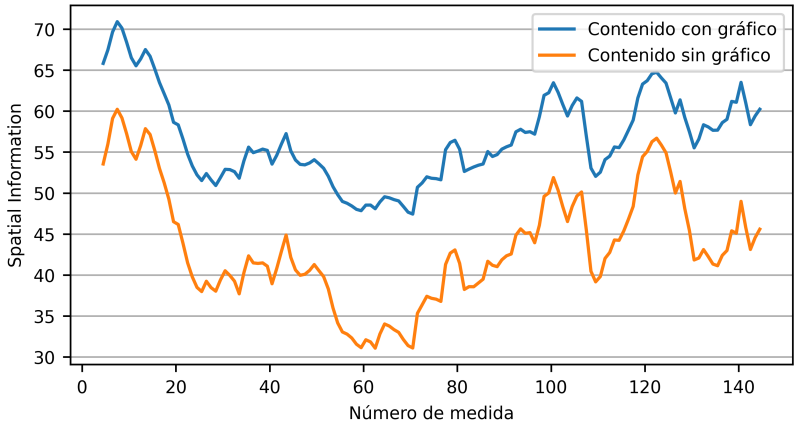


**Figura A.1:** Influencia de grafismos en la métrica *Spatial Information*.

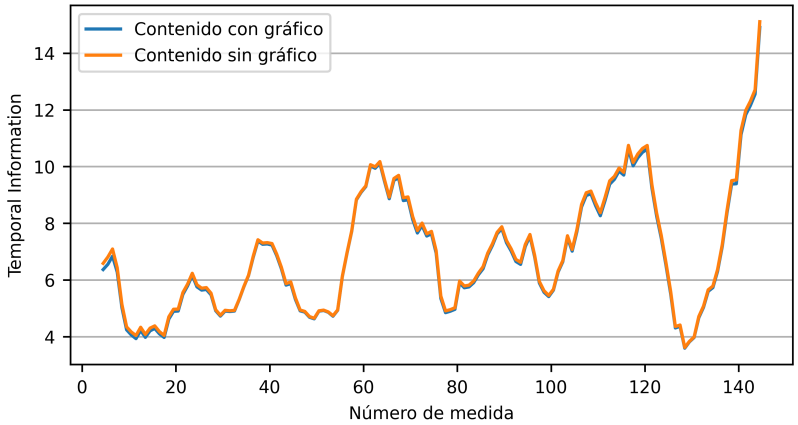


**Figura A.2:** Influencia de grafismos en la métrica *Temporal Information*.

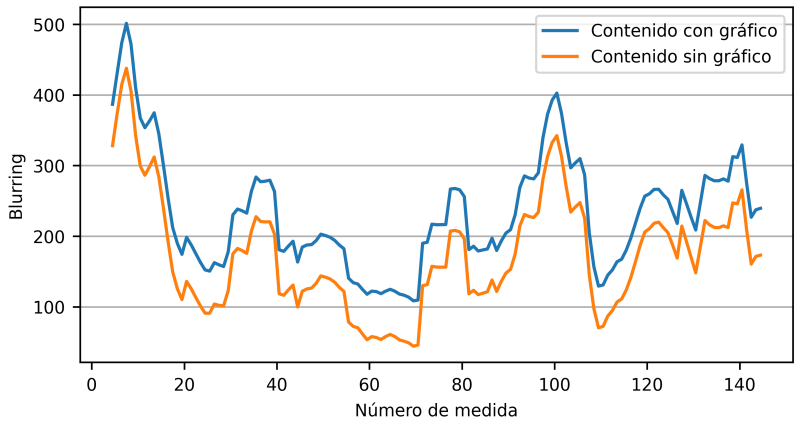


**Figura A.3:** Influencia de grafismos en la métrica *Blurring*.

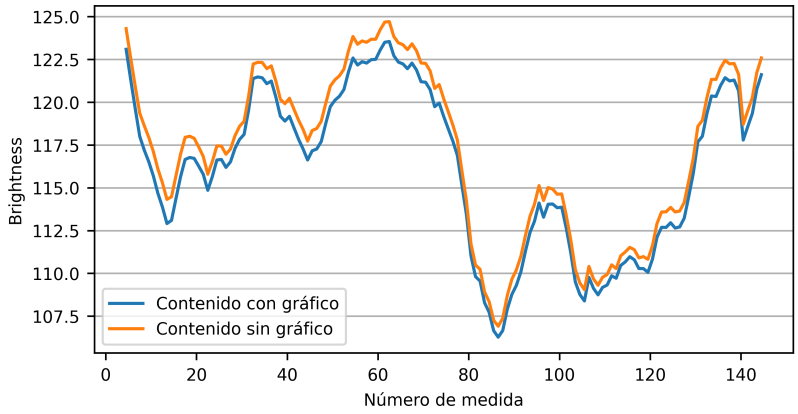


**Figura A.4:** Influencia de grafismos en la métrica *Brightness*.

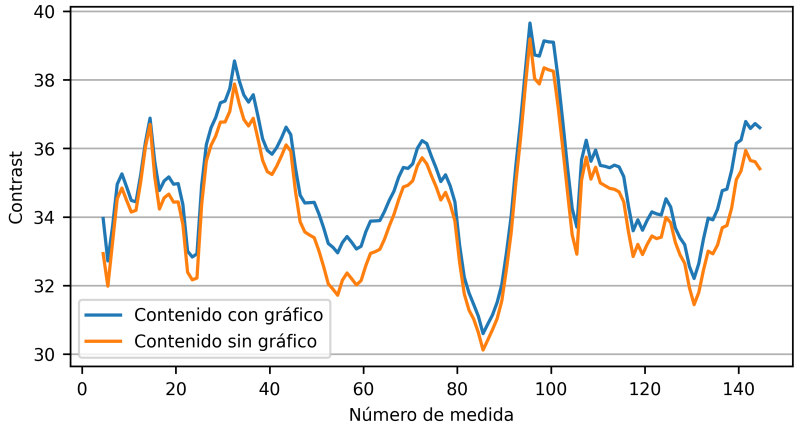


**Figura A.5:** Influencia de grafismos en la métrica *Contrast*.

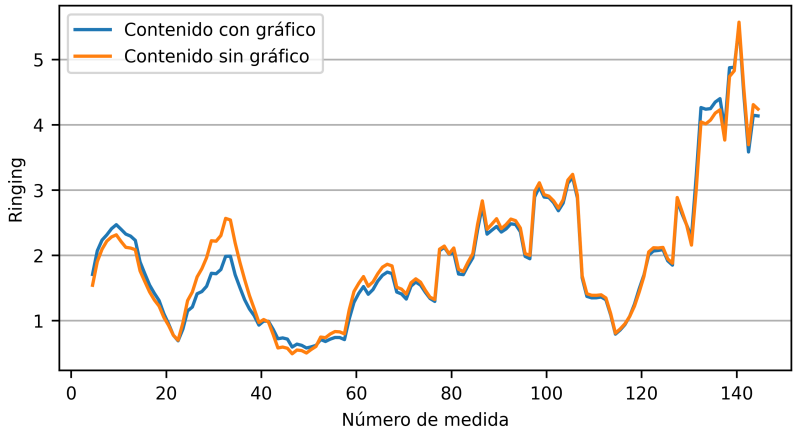


**Figura A.6:** Influencia de grafismos en la métrica *Ringing*.

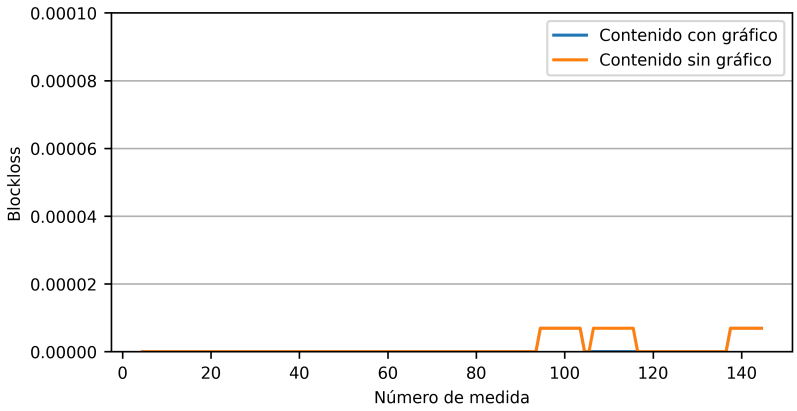


**Figura A.7:** Influencia de grafismos en la métrica *Blockloss*.

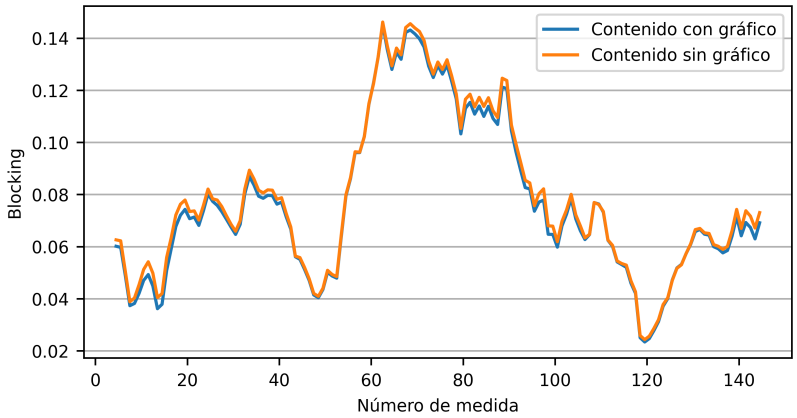


**Figura A.8:** Influencia de grafismos en la métrica *Blocking*.

# Apéndice B

# Influencia del uso de promedios en las métricas de evaluación de calidad

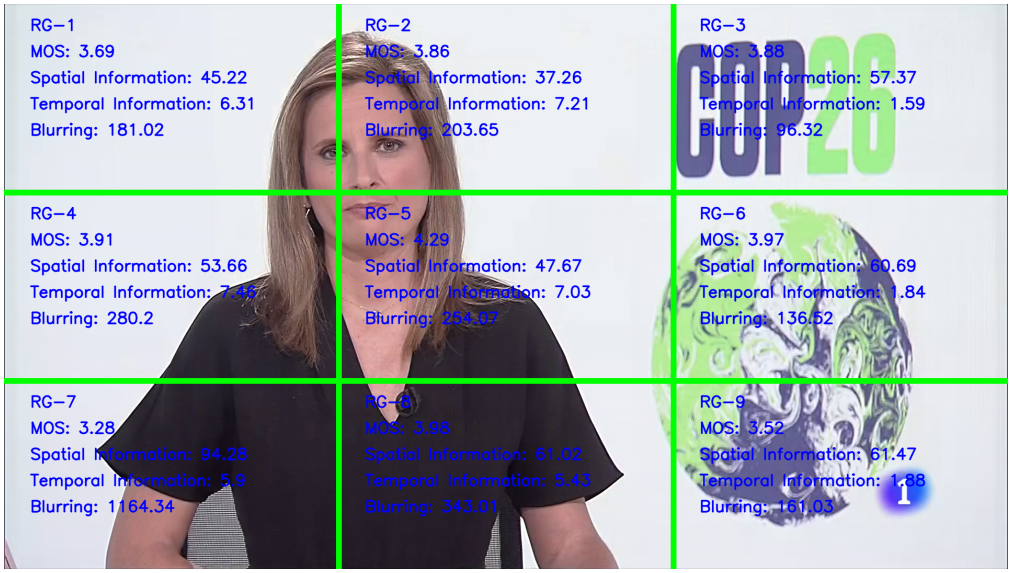


**Figura B.1:** Influencia del uso de promedios en una secuencia de prueba de La1 HD.

**Tabla B.1:** Influencia del uso de promedios. Extracción de características en una secuencia de prueba de La1 HD.

| **Métrica de evaluación** | **Número de región** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **RG-1** | **RG-2** | **RG-3** | **RG-4** | **RG-5** | **RG-6** | **RG-7** | **RG-8** | **RG-9** |
| *Spatial Information* | 45.22 | 37.26 | 57.38 | 53.66 | 47.67 | 60.69 | 94.28 | 61.02 | 61.47 |
| *Temporal Information* | 6.31 | 7.21 | 1.59 | 7.46 | 7.03 | 1.84 | 5.90 | 5.43 | 1.88 |
| *Blurring* | 181.02 | 203.65 | 96.32 | 280.20 | 254.07 | 136.52 | 1164.34 | 343.01 | 161.03 |
| *Brightness* | 123.47 | 123.82 | 122.35 | 117.84 | 118.27 | 120.55 | 105.09 | 101.24 | 118.82 |
| *Contrast* | 18.65 | 18.77 | 25.76 | 26.89 | 26.61 | 22.08 | 29.36 | 26.72 | 26.17 |
| *Ringing* | 7.25 | 2.73 | 0.00 | 6.72 | 4.68 | 0.00 | 6.33 | 19.39 | 0.64 |
| *Blockloss* | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| *Blocking* | 0.12 | 0.08 | 0.26 | 0.11 | 0.20 | 0.37 | 0.12 | 0.15 | 0.13 |

**Tabla B.2:** Influencia del uso de promedios. Error en la extracción de características en una secuencia de prueba de La1 HD.

| **Métrica de evaluación** | **Valor promedio** | **Valor modo normal** | **Error (MAE, en %)** |
|---|---|---|---|
| *Spatial Information* | 57.63 | 60.57 | 3.35 |
| *Temporal Information* | 4.96 | 5.81 | 3.00 |
| *Blurring* | 313.35 | 313.07 | 0.04 |
| *Brightness* | 116.83 | 116.83 | 0.00 |
| *Contrast* | 24.56 | 30.21 | 15.38 |
| *Ringing* | 5.30 | 3.77 | 16.93 |
| *Blockloss* | 0.00 | 0.00 | 0.00 |
| *Blocking* | 0.17 | 0.17 | 0.47 |
| **Vector de caract.** | **-** | **-** | **4.90** |

# Apéndice C

# Método propuesto de detección de saliencia


**Figura C.1:** Primera imagen de una secuencia de prueba de La1 HD.


**Figura C.2:** Segunda imagen de una secuencia de prueba de La1 HD.


**Figura C.3:** Detección de caras en la imagen de una secuencia de prueba de La1 HD.

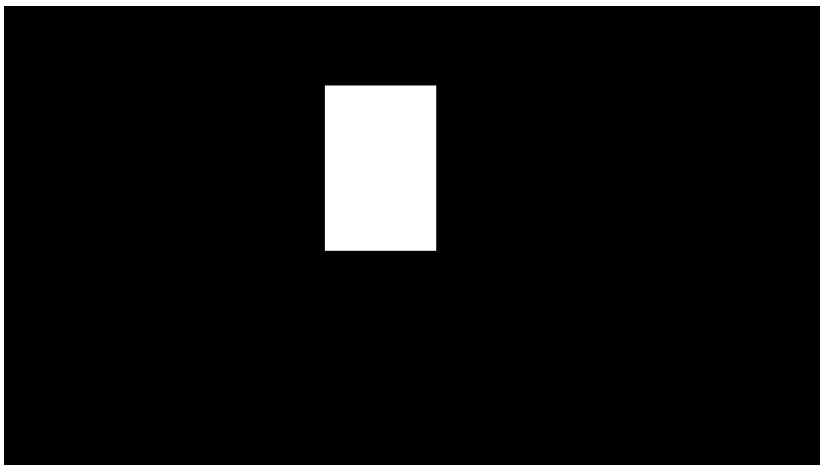

**Figura C.4:** Mapa de saliencia con detección de caras.

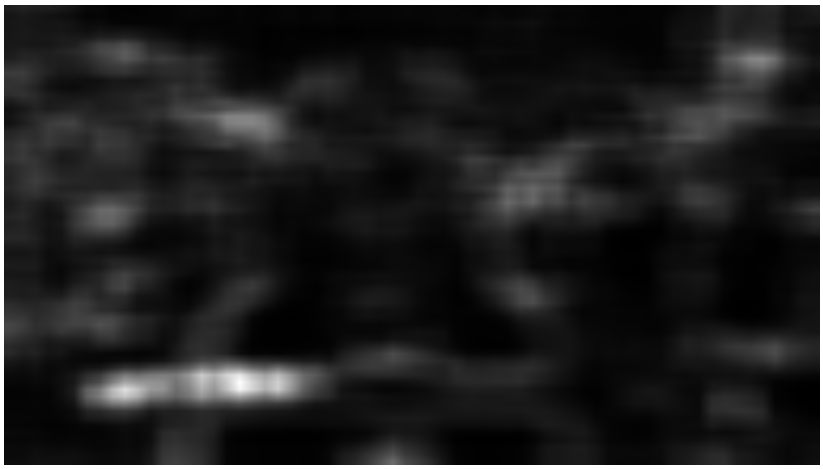

**Figura C.5:** Mapa de saliencia estática con el método *Spectral Residual*.


**Figura C.6:** Mapa de saliencia estática con el método *Fine Grained*.

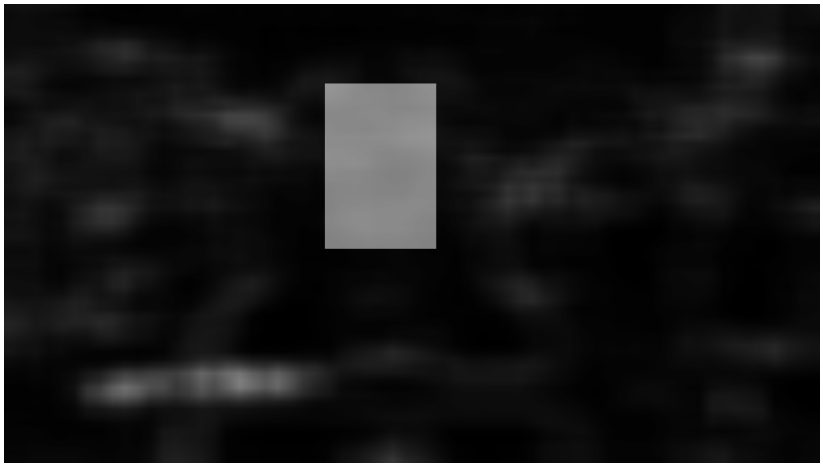

**Figura C.7:** Mapa de saliencia espacial con el método *Spectral Residual*.

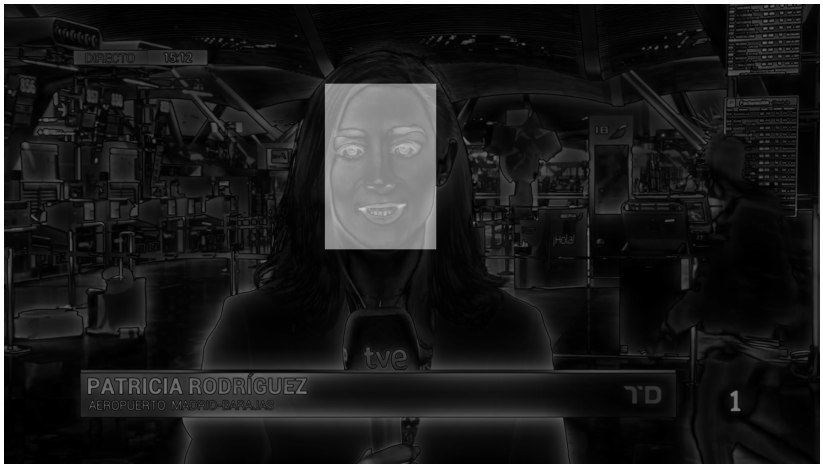


**Figura C.8:** Mapa de saliencia espacial con el método *Fine Grained*.

**Figura C.9:** Mapa de saliencia temporal.

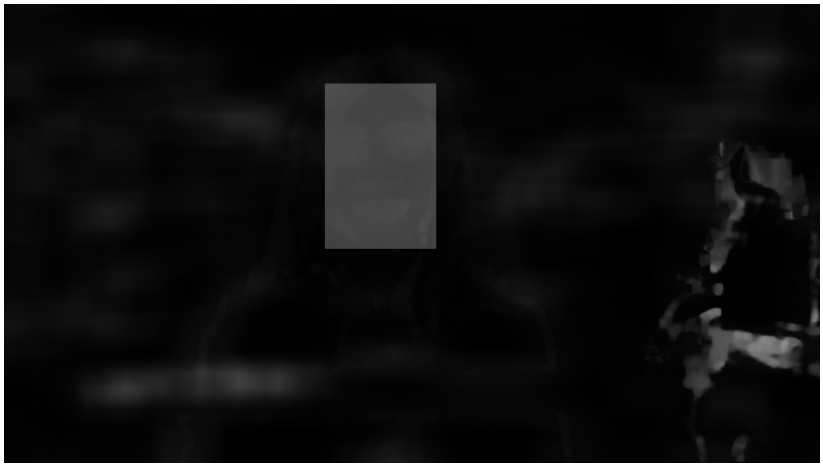

**Figura C.10:** Mapa de saliencia espacio-temporal con método SR y 50 %-50 %.

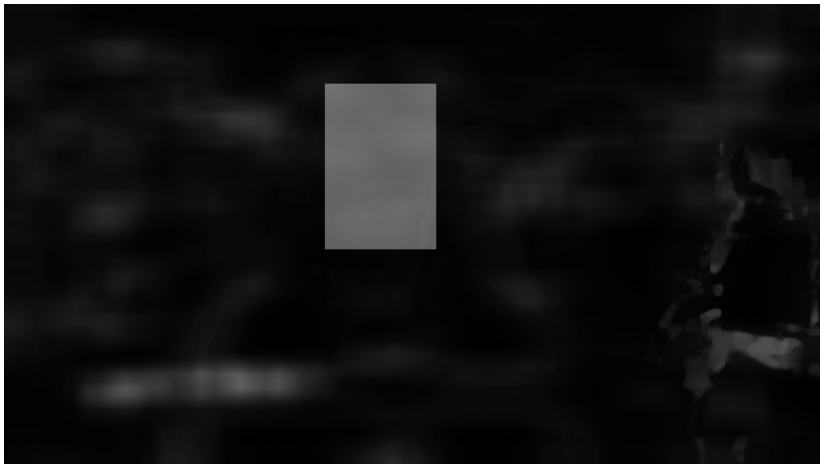

**Figura C.11:** Mapa de saliencia espacio-temporal con método SR y 75 %-25 %.