



Plataforma modular para la codificación y distribución interactiva de contenidos VR360 basada en campo de visión

Miguel Fernández-Dasi¹, Miguel A. Torres-Font², Mario Montagud^{1,2} y Miguel Garcia-Pineda².

¹Media & Internet Area

²Departamento de Informática

Fundación i2CAT

Universitat de València

C/ Gran Capità 2-4 Edifici Nexus I, Barcelona

Av. De la Universitat, s/n. Burjassot. Valencia

miguel.fernandez@i2cat.net, mitofont@alumni.uv.es, mario.montagud@i2cat.net, miguel.garcia-pineda@uv.es

El consumo de contenidos multimedia ha aumentado mucho en los últimos años. Ello incluye especialmente a los sistemas que ofrecen contenidos inmersivos, como el vídeo VR360, ya que ofrecen experiencias más realistas a los usuarios finales. Pero estos sistemas demandan unos requisitos de red, procesamiento, etc. que suponen desafíos desde el punto de vista técnico siempre y cuando se quiera ofrecer contenidos de alta calidad. Para superar estos retos y limitaciones en este artículo se presenta una plataforma modular extremo-a-extremo para la captura, aplicación de varias estrategias de proyección, trans-codificación, conversión a Dynamic Adaptive Streaming over HTTP (DASH), distribución de baja latencia y consumo interactivo de vídeos VR360, así como para la medida de una gran variedad de métricas Quality of Service (QoS) / Quality of Experience (QoE) de interés. La plataforma constituye un testbed idóneo para investigar sobre ciertos aspectos y procesos esenciales a lo largo de la cadena de distribución, como son la codificación y distribución de los contenidos, teniendo en cuenta potenciales regiones de interés (Regions of Interest (RoI)) en los vídeos y los campos de visión (Field of View (FoV)) de los dispositivos de consumo utilizados.

Palabras Clave- VR360, DASH, Region of Interest (RoI), Field of View (FoV).

I. INTRODUCCIÓN

La producción y consumo de contenidos multimedia se han incrementado progresivamente en los últimos años. Asimismo, la creciente resolución de dichos contenidos, así como su necesidad de distribuirlos en entornos heterogéneos (ej. redes y dispositivos de consumo con capacidades muy variables), no sólo requieren la disponibilidad de anchos de banda y capacidades de procesado suficientes, sino también de técnicas de codificación y distribución adaptativa (ej. basadas en streaming adaptativo HTTP). Estos requisitos son, a su vez, más críticos cuando se refiere a contenidos de

Realidad Virtual (RV), tales como el vídeo 360° (en adelante VR360), pues capturan mayor cantidad de información y suelen tener resoluciones altas con el fin de proporcionar un nivel de inmersión satisfactorio.

La comunidad científica sigue muy activa con tal de superar retos y limitaciones existentes para proporcionar de manera más eficiente servicios distribuidos de vídeo VR360 [1, 2]. Este artículo se centra en esta temática, presentando una plataforma extremo-a-extremo modular que posibilita un banco de pruebas de investigación sobre aspectos esenciales como son la captura, procesado, distribución y consumo interactivo de vídeo VR360. En particular, este artículo presenta evidencias preliminares sobre los beneficios que pueden aportar el diseño y adopción de soluciones innovadoras y adaptativas basadas en el campo de visión (*Field of View* (FoV)) y/o regiones de interés (*Region of Interest* (RoI)), en cuanto a el sistema de proyección 360° (ej. *Equirectangular Projection* (ERP) vs *Cubemap Projection* (CMP)) y a la codificación de vídeo. Estas soluciones se basan en la premisa de que no toda la información capturada en el panorama 360° tiene la misma relevancia, así como que los usuarios sólo pueden visualizar en todo momento una porción del espacio 360°, determinado por el FoV de su dispositivo de consumo, que puede oscilar entre 90°-130°. Además, la plataforma se complementa con una serie de herramientas para el registro de métricas de *Quality of Service* (QoS) indicadoras no sólo del consumo de recursos de cada estrategia y componente bajo análisis, sino también de su impacto sobre la *Quality of Experience* (QoE) percibida.

[1] <https://zenodo.org/record/3909281#.YMz-TmgzaiN>

II. TRABAJOS RELACIONADOS

En los últimos años la comunidad científica y la industria han dedicado esfuerzos considerables en el área referente a sistemas de vídeo VR360 streaming, proponiendo optimizaciones a lo largo de la cadena extremo-a-extremo [1, 2], incluyendo diversas alternativas en cuanto a estrategias de proyección y codificación basadas en FoV / RoI. A continuación, se destacan brevemente algunas de las contribuciones más relevantes en el área, relacionadas con el trabajo que se presenta. En [3] se presenta una plataforma extremo-a-extremo VR360 que soporta ERP y CMP, así como la medida de métricas, pero no soporta codificación basada en RoI. En [4] se presenta una solución basada en proyección ERP que permite concentrar la calidad en la RoI, pero requiriendo una superposición de tramas. En [5] se propone concentrar la resolución en una determinada RoI cuando se usa CMP, pero utilizando diferentes flujos de vídeo. En [6] se propone una solución para concentrar la calidad dinámicamente en la RoI cuando se utiliza CMP, mediante estimaciones de proyecciones para cada posible dirección/orientación de la cámara virtual. También se han tratado de proponer alternativas más avanzadas que ERP y CMP, como es *Equi-Angular Cubemap* (EAC) [7]. EAC ha sido adoptada por Youtube y proporciona mejoras en cuanto a la uniformidad de densidad de píxeles, aunque requiere una mayor complejidad computacional.

En general, las soluciones existentes se basan en el uso de códecs con restricciones de licencia, requieren el uso de varios codificadores y/o flujos (ej. codificación escalable, multi-nivel, etc.), y presentan problemas de compatibilidad con navegadores web (ej. uso de H.265). En este trabajo se

persiguen soluciones más simples, pero efectivas, basadas en el uso de códecs tradicionales como H.264, que son compatibles con los navegadores web, y en procesos de codificación y flujos multimedia únicos. Además, se persiguen soluciones que impliquen modificaciones y cargas mínimas en el reproductor. Ello no sólo permitirá una mayor escalabilidad e interoperabilidad, sino también un mejor rendimiento en sistemas interactivos.

III. PLATAFORMA VR360

En esta sección se presenta la plataforma VR360 desarrollada (ver Fig. 1), con sus bloques / módulos principales: preparación de contenidos (CP), servidor de contenidos (CS), señalización y distribución de contenidos (CD) y consumo de contenidos (CC).

A. Preparación de contenidos – CP

La plataforma puede tomar como entrada vídeos 360° almacenados o en vivo. En cuanto a la proyección de los vídeos VR360, no sólo se soporta las proyecciones típicas ERP y CMP uniformes (parte superior de Fig.2), sino que se han propuesto soluciones innovadoras para concentrar la resolución en determinadas RoI (parte inferior de Fig.2 para CMP y Fig.3 para ERP). En el caso de CMP, se propone un re-ajuste del cubo para, con una determinada relación de aspecto, concentrar la resolución en una o varias caras. En el caso de ERP, se pueden aplicar dinámicamente diferentes ajustes de codificación a la región exterior al RoI (Fig.3) gracias al desarrollo de un sistema y algoritmo de codificación interactivo y adaptativo, alimentado a partir de patrones de visionado enviados desde el reproductor multimedia.

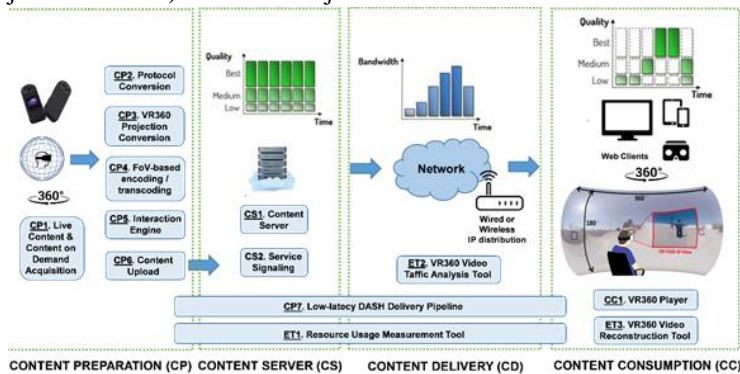


Fig. 1. Diagrama de la plataforma VR360 desarrollada

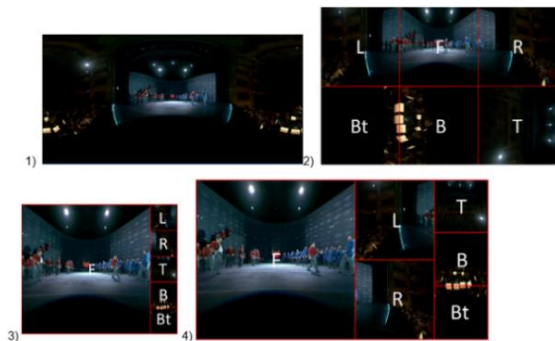


Fig. 2. Proyecciones VR360 consideradas: 1) ERP; 2) CMP 3:2; 3) CMP 6:5 concentrando la resolución en la cara frontal (F); 4) CMP 11:6 concentrando la resolución en el panorama horizontal

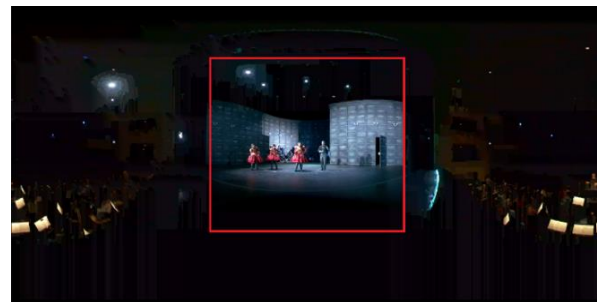


Fig. 3. Codificación basada en RoI aplicada a proyección ERP



A continuación, el flujo de vídeo se codifica y se convierte a Dynamic Adaptive Streaming over HTTP (DASH) [7, 8], con los ajustes deseados, e integrando sus recientes funcionalidades para su distribución en baja latencia. En el caso de recepción de vídeos en vivo, este bloque también soporta la conversión entre protocolos origen y destino, así como la trans-codificación del flujo. Finalmente, el vídeo resultante se transfiere a un servidor (web) de contenidos (ej. Apache).

Las soluciones propuestas no están restringidas al uso de códecs avanzados que impliquen problemas de licencias y/o compatibilidad en ciertas plataformas y/o navegadores web, ni requieren el uso de varios flujos de transmisión, aportando así ventajas significativas para su adopción.

B. Señalización y Distribución de contenidos – CS y CD

Además de los contenidos, el servidor web aloja un fichero para la indexación de los mismos, incluyendo información relevante sobre los mismos (ej. título, duración, tipo de proyección...).

C. Consumo de contenidos – CC

En cuanto al consumo de contenidos, se ha adoptado un reproductor web *open-source* [7] y extendido con tal de: 1) procesar las proyecciones basadas en RoI propuestas (a partir de la información proporcionada por el fichero de indexación); 2) detectar y reportar periódicamente al sistema de codificación los patrones de visionado (centro del FoV), cuando se utiliza la proyección ERP basada en RoI. Una ventaja fundamental de las soluciones basadas en RoI propuestas en este trabajo es que no requieren modificaciones en el reproductor en caso de adoptar ajustes de RoI pre-establecidos (ej. caras del cubo) o mínimas en el caso de requerir ajustes de RoI dinámicos, simplemente teniendo que reportar el centro del FoV a través de una conexión vía sockets con el sistema de codificación del bloque de preparación de contenidos.

IV. Herramientas de Medida de Métricas

La plataforma se complementa con una serie de herramientas para la medida de métricas QoS/QoE, con tal de determinar los beneficios de las soluciones propuestas.

A. Medida de Consumo de Recursos

Herramienta software *open-source* [9] que permite medir el uso de CPU (%), GPU (%) y memoria RAM (MB) para procesos de Windows, haciendo uso de llamadas a la *Powershell*. Puede ser aplicada a cada componente de la plataforma, aunque resulta de especial interés para los módulos de (trans-)codificación y consumo de contenidos.

B. Análisis de tráfico DASH

Herramienta software *open-source* [9] para capturar y analizar el tráfico DASH, así como las peticiones HTTP asociadas, para una sesión de streaming determinada,

haciendo uso de *tshark* (versión de consola de *WireShark*). Esta herramienta no sólo captura el fichero *Media Presentation Description* (MPD) asociado al vídeo DASH seleccionado, sino también los parámetros asociados a cada segmento DASH solicitado/descargado (URL, códec, resolución, *bitrate*, *fps*...), así como su tamaño, número de segmentos asociados y retardo de ida y vuelta .

C. Medida de Patrones de Visionado VR360

A partir de las medidas periódicas del centro del FoV en el reproductor, se ha desarrollado un script para representar el mapa de calor de las regiones / patrones de visionado en un plano 2D, en forma rectangular (para ERP) o de T tumbada (para CMP).

V. TESTS Y RESULTADOS PRELIMINARES

En esta sección se presentan brevemente algunos de los tests realizados, así como los resultados preliminares obtenidos, cuando: 1) se compara la proyección ERP con CMP, y sus variantes; y 2) se analiza el potencial impacto de aplicar codificación basada en RoI para ERP. La capacidad de la plataforma para registrar métricas QoS, tales como la calidad DASH seleccionada, el *throughput* estimado o el nivel de ocupación del buffer de reproducción, se demostró en [3].

A. Comparativa entre ERP y (variantes de) CMP

En primer lugar, se codificaron 3 vídeos DASH con los mismos niveles de calidad (resoluciones y bitrates) equivalentes, en función de sus relaciones de aspecto, para comparar tanto a nivel QoS como QoE las siguientes condiciones de test (TC): *TC1*) ERP con CMP; *TC2*) CMP 3:2 uniforme con CMP 6:5; y *TC3*) CMP 3:2 uniforme con CMP 11:6. En cuanto a eficiencia de compresión, los resultados confirman que se puede conseguir un ahorro de tamaño de vídeo y, por tanto, de capacidad en servidor y de ancho de banda, entre 20-25% cuando se usa CMP con respecto al uso de ERP, ambos en sus variantes uniformes. Estos resultados están en línea con los obtenidos en el estado del arte [1, 2]. A su vez, las variantes de CMP basadas en RoI propuestas introducen una ganancia con respecto a CMP uniforme entre el 10-15% de tamaño.

Además, se realizaron pruebas subjetivas con usuarios ($N=24$ participantes) comparando las tres condiciones de test, alternando el orden de presentación de los vídeos a consumir. En dichas pruebas, se presentaron cuestionarios sobre percepción QoE (utilizando medidas *Mean Opinion Score*) e inmersión (*Igroup Presence Questionnaire (IPQ)* [10]). De manera interesante, los resultados indican que a pesar de la reducción de tamaño (y por tanto ahorro de recursos y costes) el uso de CMP no tiene un impacto negativo sobre los niveles de QoE e inmersión percibidos, así como que el uso de las variantes CMP basadas en RoI pueden incluso aportar mejoras, en función del vídeo.

Finalmente, la Fig. 4 representa el mapa de calor de los patrones de visionado agregados de los 24 usuarios para el mismo vídeo, codificado en proyección ERP y CMP. Aunque se trate de un vídeo con escenas relevantes a lo largo del entorno 360°, el visionado se concentra en regiones específicas, vislumbrando los potenciales beneficios del procesado y distribución basado en RoI.

B. Uso de codificación basada en RoI para ERP

En segundo lugar, se realizaron unas iteraciones de tests para analizar preliminarmente el impacto de aplicar en un vídeo determinado (resolución 1280x720, duración 1'30'') diferentes valores de Constant rate factor (CRF) para las regiones exteriores a una RoI determinada (posición fija y tamaño variable en cada iteración de test: 1/4, 1/3 y 1/2 de la resolución del vídeo), en cuanto al tamaño del vídeo y a métricas objetivas de calidad, como Structural Similarity Index (SSIM). Los resultados se pueden observar en la Fig. 5.

Por un lado, en términos de tamaño, el uso de la RoI puede resultar en una reducción de hasta un 500% manteniendo índices de calidad prometedores. En la Fig. 5 se puede observar que a medida que se disminuye el tamaño de RoI, también disminuye el tamaño del vídeo, ya que existe una mayor región de vídeo con unos niveles de calidad inferiores. El tamaño de la RoI dependerá del tipo del dispositivo de consumo utilizado, del tipo de vídeo, e incluso podría ajustarse en base a los patrones de visionado. Por otro lado, las gráficas de la Fig. 5 muestran que para rangos de CRF hasta alrededor de 27, el valor de SSIM obtenido se acerca a la unidad, consiguiendo un ahorro de tamaño en torno al 200%. Estos resultados reflejan claramente el potencial beneficio de aplicar codificación basada en la RoI, especialmente en el caso de vídeos VR360 en los que las regiones exteriores a la RoI se pueden ajustar para que se encuadren fuera del FoV.

VI. CONCLUSIONES Y TRABAJOS FUTUROS

Este artículo ha presentado una plataforma modular extremo-a-extremo para la captura, aplicación de varias estrategias de proyección, trans-codificación, conversión a DASH, distribución y consumo interactivo de vídeos VR360, así como para la medida de métricas QoS / QoE. La plataforma constituye un *testbed* idóneo para investigar sobre ciertos aspectos y procesos esenciales, como son la codificación y distribución de los contenidos, teniendo en cuenta potenciales RoI y FoV. Se han presentado evidencias preliminares sobre los potenciales beneficios que aportan este tipo de estrategias.

Como trabajo futuro, se va a investigar sobre servicios de streaming VR360 aplicando este tipo de técnicas y estrategias basadas en RoI / FoV para una gran variedad de vídeos y escenarios, así como se evaluarán sus prestaciones en cuanto a latencia y escalabilidad. Finalmente, se especificarán estrategias de modelado QoE para servicios de VR360 streaming basados en DASH [8].

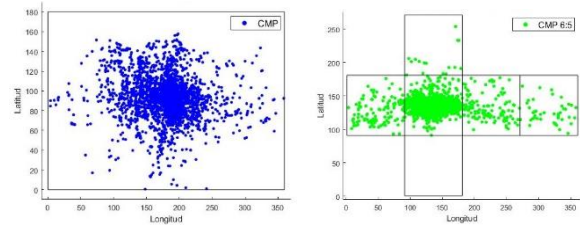


Fig. 4. Mapa de calor de proyección ERP (izq) y CMP (der).

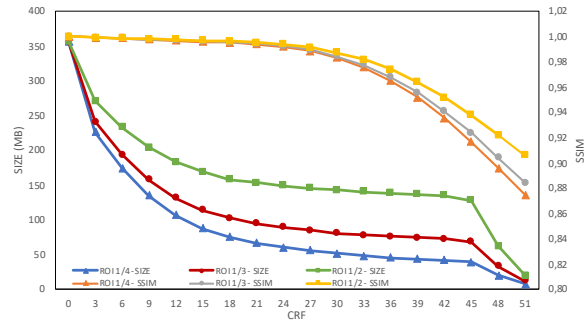


Fig. 5. Análisis de tamaños en MB frente al valor CRF utilizado en codificación de vídeo.

AGRADECIMIENTOS

Este trabajo ha sido financiado parcialmente por la Generalitat Valenciana en el marco del proyecto GV/2020/052, por la Comisión Europea en el marco del proyecto EU H2020 Respond-A (ID 883371), por la red temática Everest (RED2018-102383-T) y por la UV (UV-INV-AE-1564749). Asimismo, el trabajo de Miguel Fernández ha sido financiado por la Secretaria d'Universitats i Recerca de la Generalitat de Catalunya i del Fons Social Europeu (Personal Novel 2021 FI_B_01041) y el de Mario Montagud por una Beca postdoc JdC-Incorporación (MICINN, IJCI-2017-34611).

REFERENCIAS

- [1] C.L. Fan, W.C. Lo, Y.T. Pai, C.H. Hsu, "Survey on 360° Video Streaming: Acquisition, Transmission, and Display", *ACM Comput. Surv.* 52, 4, Article 71, Sept. 2019, 36 pages
- [2] M. Xu, C. Li, S. Zhang, P.L. Callet, "State-of-the-Art in 360° Video/Image Processing: Perception, Assessment and Compression", *IEEE JSAC*, vol. 14, no. 1, pp. 5-26, Jan. 2020
- [3] M. Montagud, E. Meyerson, I. Fraile, S. Fernández, "Modular Testbed for KPI Monitoring in Omnidirectional Video Streaming Scenarios", *MMEDIA 2019*, València (Spain), March 2019
- [4] A. Xu, X. Chen, Y. Liu and Y. Wang, "A Flexible Viewport-Adaptive Processing Mechanism for Real-Time VR Video Transmission", *IEEE ICMEW*, 2019.
- [5] D. Gómez, J. A. Núñez, I. Fraile, M. Montagud, S. Fernández, "TiCMP: A lightweight and efficient Tiled Cubemap projection strategy for Immersive Videos in Web-based players", *ACM NOSSDAV'18*, Amsterdam, June 2018
- [6] P. Yu, Timokhin, M.V. Mikhaylyuk, E.M. Vozhegov, "Efficient methods and algorithms to synthesize 360-degree video based on cubemap projection of virtual environment", *ISP RAS*, 32:4, 2020
- [7] J. Lin, et al., "Efficient Projection and Coding Tools for 360° Video", *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 84-97, March 2019
- [8] M. Garcia-Pineda, J. Segura-García; S. Felici-Castell; A. Soriano-Asensi M. Montagud, "Modelando la QoE en entornos DASH", *JITEL 2019*, Zaragoza (Spain), October 2019
- [9] M. Montagud, J. Antonio De Rus, R. Fayos-Jordán, M. Garcia-Pineda J. Segura-García, "Open-Source Software Tools for Measuring Resources Consumption and DASH Metrics", *ACM MMSYS 2020*, Istanbul (Turkey), June 2020
- [10] Igroup Presence Questionnaire (IPQ), <http://www.igroup.org/pq/ipq/index.php> Acceso en 30/09/2021

