

Adaptive Strategies For Enhancing Quality of Experience in 360° Video Streaming

Ioanna Valiandi
VIDEOMICS Group
CYENS Centre of Excellence
Nicosia, Cyprus
i.valiandi@cyens.org.cy

Marios S. Pattichis
Department of Electrical and
Computer Engineering
University of New Mexico
Albuquerque, NM 87131-0001,
USA
pattichi@unm.edu

Efthymou Kyriacou
Department of Electrical Eng.,
Computer Engineering and
Informatics,
Cyprus University of Technology
Limassol, Cyprus
efthymou.kyriacou@cut.ac.cy

Andreas S. Panayides
VIDEOMICS Group
CYENS Centre of Excellence
Nicosia, Cyprus
a.panayides@cyens.org.cy

Abstract— The unprecedented growth of immersive virtual and mixed reality technologies is driving the increasing demand of 360° video streaming applications. Facilitating unique opportunities, but also challenges, 360° videos require substantially higher video resolutions, and hence bandwidth demands, for a better viewing experience. This study examined four tile-based Region of Interest (ROI) encoding techniques to optimize 360° video encoding and streaming. These techniques encode the user's viewport regions with higher quality compared to the background, non-ROI areas. A dataset of ten 4K omnidirectional videos at 30 frames per second was used to evaluate the performance of four distinct tile-based ROI methods using four video encoding standards, namely hevc_nvenc, x265, h264_nvenc, and x264. Overall, the proposed tile-based ROI encoding methods achieve substantial bitrate reductions compared to the uniform full-frame encoding while maintaining adequate subjective video quality levels, demonstrating their potential for efficient and scalable 360° video streaming. The ROI-3 method, which uses a sliding window approach to encode the user's field of view based on eye gaze movement (EM) data coordinates translated to tile regions, was found to achieve the best trade-off between bitrate demands savings and perceptual quality for all investigated video compression standards.

Keywords—360° videos, Video compression, HEVC, AVC, 360° Video Streaming, Video Tiling, Region Of Interest

I. INTRODUCTION

The advent of 360° video technology has revolutionized immersive media experiences, enabling users to interact with immersive, panoramic environments [1],[2],[3]. As 360° video content becomes increasingly popular across various platforms [4],[5],[6], there is an urgent need to address the significant challenges posed by the high data demands and bandwidth requirements associated with streaming and delivering such videos [7]. Given the large file sizes and high bitrate demands of 360° videos, particularly at higher resolutions such as 4K and beyond, traditional video encoding methods often struggle to meet the needs of real-time streaming while maintaining acceptable video quality. The latter frequently results in excessive data usage, reduced video

quality, or poor user experience due to limited computational and bandwidth resources [8],[9],[10].

Traditional video encoding standards for 2D videos are also applied to 360° videos by projecting them into 2D formats like the Equirectangular Projection (ERP). Tiling mechanisms partition ERP videos into grids of rectangular tiles, which are individually encoded to enhance streaming performance. However, fixed tiling schemes often fail to align with users' fields of view (FoV), resulting in unnecessary quality and bandwidth demands overhead, particularly in polar regions that are less visible. Region of Interest (ROI) encoding techniques have emerged as a promising solution to overcome these limitations. They encode selective areas of the video, particularly those in the user's field of view, at higher quality while aggressively compressing less important regions. This selective encoding not only optimizes video quality but also significantly reduces the overall bitrate, making streaming more feasible, even in bandwidth-constrained environments [9],[11],[12]. Thus, determining the optimal number of tiles is critical towards a balanced video quality and bitrate demands trade-off [13],[14].

This study proposes and evaluates four different strategies to identify the ROIs in a video that correspond to the user's field of view (FoV). It provides a systematic comparison of these methods to determine the most effective strategy for optimizing video streaming performance. By comparing ROI identification methods and video encoding standards (H.264/AVC and H.265/HEVC), this work seeks to highlight the trade-offs between bitrate savings and subjective and objective video quality. It further examines the impact of ROI size on video quality metrics, demonstrating significant bitrate demands reductions while preserving perceptual quality. Overall, the study's findings aspire to provide insights into implementing tile-based variable quality encoding ROI techniques for immersive media, paving the way for optimized 360° video streaming in network-constrained environments.



Figure 1. Examples of video frames captured from the 360° dataset.

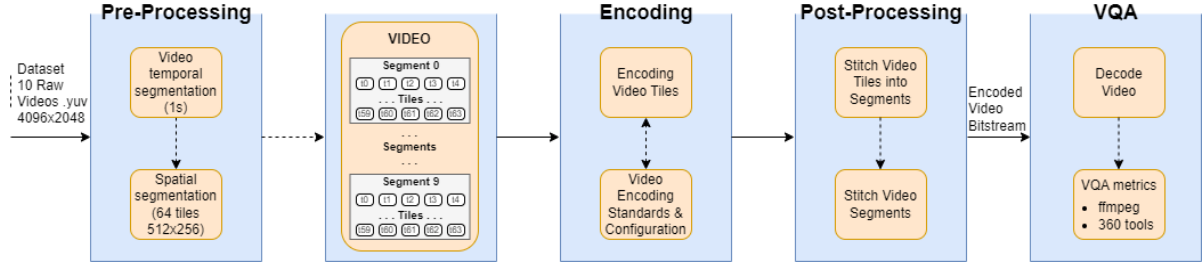


Figure 2. ROI Tile-Based Encoding Overview

A dataset of ten 360° omnidirectional videos with 4K resolution (4096x2048 pixels@30fps) was utilized. During the pre-processing phase, temporal and spatial segmentation were implemented, where each video was organized into 10 video segments, each lasting one second and then split them into 64 distinct tiles (512x256 pixels). Each tile was encoded using the selected encoding parameters based on prioritization in the ROI regions. The post-processing phase involved stitching all segments and tiles together to reconstruct the entire video. In the video quality assessment (VQA) phase, metrics from the 360tools capable of handling 360° videos were used to evaluate the quality of the compressed videos.

Table I. Tile-Based Region of Interest (ROI) computation methods

ROI Method	Details	SHAPE (8x8)
ROI = 0	1 Tile (for each user)	
ROI = 1	9 Tiles (cross)	
ROI = 2	9 Tiles (square 3x3)	
ROI = 3	X Tiles (Square 3x3 Filtering)	

II. METHODOLOGY

A. 360° Video Dataset

The video dataset used in the present study consists of ten omnidirectional 360° videos [15]. All the videos have diverse content, three of them comprising relatively high motion scenes, while the others being mostly static. The videos are in YUV, equirectangular projection (ERP) format and have a 4K resolution (4096x2048 pixels) at 30 frames per second and a duration of 10 seconds each. The names of the videos are (a) Xiao Guang, (b) Pagoda, (c) Football Match, (d) Concert Live, (e) Boat in Park, (f) Salon, (g) Buddha Cave, (h) Press Conference, (i) Great Wall, and (j) Shooting. Figure 1 illustrates examples of the first four videos. Additionally, the dataset provides publicly available traces, capturing the eye gaze movement (EM) and head movement (HM) of users watching the videos using a head-mounted display (HMD).

B. Region of Interest (ROI) Definition

The proposed methods necessitate FoV definition and subsequent mapping to tile-based ROI for variable quality encoding. We processed the publicly available traces, capturing the eye gaze movement (EM) data consisting of (x,y) coordinates sampled at a frequency of 60 samples per second, for 20 individual users. To simplify the analysis, we computed the average of the 60 samples within each segment (i.e., two times per frame), reducing the data to a single representative coordinate per segment. These averaged (x,y)

coordinates were then mapped to their corresponding tiles using the Euclidean distance metric, ensuring precise spatial localization.

The ROI was then determined for each of the four proposed ROI computation methods, as depicted in Table I. The first method, referred to as ROI=0, assigns only a single tile as the ROI, corresponding to the user's exact gaze point. Typically, this approach captures an ROI of fewer than 20 tiles, as we only record one tile per user, and several users are projected to focus on the same tile. The second technique, ROI=1 expands the ROI to include a cross-shaped region comprising nine tiles. For each user, the central tile, defined in ROI=0, is used along with two adjacent tiles on the right, left, top, and bottom. The third method, ROI=2, defines the ROI as a square of nine tiles. It centers the user's gaze tile from ROI=0 and includes all neighboring tiles within the 3x3 square boundary. The final technique, ROI=3, leverages a sliding window approach to propagate features across the video. This method applies additional filtering, expanding the ROI to include tiles within a processed 3x3 square. The 3x3 filter is applied iteratively across the entire frame, which consists of 8x8 tiles. The filter functions as a sliding window and is applied iteratively to the entire grid, from the upper-left to the lower-right corner. The grid, initialized with zeros, contains a single marker to represent the tile of interest, as defined in ROI=0. The filter, consisting of uniform values, is systematically overlaid on the grid. Each subregion is checked for the presence of the marker, and if detected, the corresponding values in the subregion are modified by the filter. Finally, we define the ROI by capturing the tiles where the resulting values exceed a predefined threshold of 3, thereby identifying the regions impacted by the transformation.

Each approach varies in complexity and ROI coverage, providing adaptability for examining user interaction based on specific application requirements.

C. Tile-based ROI Encoding

This study proposes a method that involves several stages for 360° video processing and encoding, as illustrated in Figure 2. Firstly, we applied temporal segmentation to the 10 videos in our dataset, organizing them into 100 distinct video segments. Each video segment had a one-second duration, ensuring a feasible viewing duration that accounts for variations in users' viewports while enabling efficient inter-frame encoding and reliable transmission for adaptive HTTP

streaming. Secondly, we partitioned each video segment of the 4K video into 8x8 tiles, each with dimensions of 512x256 pixels.

This approach ensured fine spatial segmentation for efficient processing. As described in Section II.B, we identified the tiles corresponding to the user's viewport based on specific user-provided coordinates. This mapping enabled precise targeting of the user's FoV areas for quality prioritization during the bitrate budget allocation. Each tile was encoded separately, using a constant quantization parameter (QP), with the ROI tiles being assigned a lower QP (higher quality) during encoding than non-ROI tiles. More specifically, we used a QP of 22 to encode the viewport tiles, which reflect the focus areas within the video, and a lower QP of 37 for the non-ROI tiles, which represent the background. The video tiles were encoded using the H.264/ AVC [16] and H.265/ HEVC [17] video encoding standards, with two different implementations for each standard: ffmpeg's builds of libx264 [18] and libx265 [19], as well as NVIDIA's implementations: h264_nvenc and hevc_nvenc [20]. The latter two implementations utilized NVIDIA GPUs via ffmpeg to significantly accelerate video encoding, taking advantage of the APIs provided by the NVIDIA Video Codec SDK. For all codecs, the medium preset was used to activate the main profile.

Variable quality encoding approach implements varying compression levels between the viewport area and the rest of the video, optimizing resource allocation. It ensures that visual quality is maintained where it matters most, within the user's focus area, while simultaneously reducing overall bandwidth requirements. Once the tiles were encoded, we used ffmpeg to stitch the individual tiles back together, reconstructing the whole video segments. Subsequently, the segments were stitched together to reconstruct the complete video, ensuring spatial integrity and preserving the quality adjustments made to the targeted regions. This approach ensures efficient encoding and delivery of high-quality video while considering the user's viewpoint.

D. Full-frame Encoding

To establish a reliable benchmark, we encoded the ten videos from our dataset without any temporal segmentation, using the four selected video encoding standards' instances, namely x264, x265, h264_nvenc, and hevc_nvenc. For these full-frame encodings, we applied a fixed quantization parameter (QP) of 22, while keeping all other encoding parameters consistent with those used during the tile-based ROI encoding. This ensured uniform encoding settings across all methods. This approach provided a standardized reference for comparing bandwidth demands with respect to subjective and objective video quality. Both tile-based ROI encodings and full-frame encodings were performed on a 64-bit Windows 10 (v.22 H2) machine, which featured a 12th Gen Intel(R) Core (TM) i9-12900K (16 cores, 3.20 GHz).

E. Video Quality Assessment

We used Samsung's 360tools [21] to assess the effectiveness of the strategies we investigated for 360° videos, utilizing four different video quality metrics, namely PSNR, S-PSNR, WS-PSNR, and CPP-PSNR.

To mitigate shortcomings of the PSNR metric, a prevalent alternative has emerged in the past decade that emphasizes the luma component (PSNR Y), since it more effectively represents the strength of the monochrome signal in

comparison to the chroma components PSNR (U,V) [22]. We compute the global metric $PSNR_{611} = (6 \cdot PSNR_Y + PSNR_U + PSNR_V) / 8$, which demonstrates a superior association with the perceptual quality of videos [23].

The spherical PSNR (S-PSNR) compares the pixel values of the original and distorted frames using their uniformly sampled spherical coordinates. Specifically, the algorithm uniformly samples points on a spherical surface, reprojects them onto the reference and the compressed omnidirectional videos, and then calculates the PSNR. Points located between sampling locations in the 2D plane are assigned to the nearest neighbor. The S-PSNR metric can be used to assess videos with different panoramic projections [24]. The Craster Parabolic Projection (CPP-PSNR) projects the pixels of the reference and compressed video frames to the spherical domain and maps/resamples them into the Craster Parabolic Projection format. The CPP maintains area fidelity in the spherical domain without altering spatial resolution. PSNR is subsequently calculated in the transformed domain, allowing it to remain applicable even when the reference and compressed frames have different projection formats and resolutions [25],[26]. CPP-PSNR and S-PSNR are designed for omnidirectional video and involve resampling based on a specific projection format. To address inconsistencies arising from the resampling process from representation space to observation space, the Weighted-to-Spherically-Uniform (WS-PSNR) metric was developed. WS-PSNR evaluates the quality between two omnidirectional videos in the spherical domain. The metric measures the distortion of samples in representation space by computing the PSNR on each pixel of the projected frame, then weighting the results proportionally to the corresponding projection area the pixel occupies on the sphere [27].

III. RESULTS

In this section we provide a comprehensive comparative evaluation of the variable quality tile-based ROI encodings versus the full-frame universal quality encoding for each of the examined video codecs. We used the Bjontegaard Delta (BD-Rate) metric to compute the average bitrate savings for each case. The obtained results are depicted in Table II across the four different objective video quality assessment metrics, PSNR, S-PSNR, WS-PSNR and CPP-PSNR.

Three principal observations may be inferred from the BD-Rate tables. First, the BD-Rate savings across the four metrics show a high degree of correlation. This consistency indicates that the variable quality tile-based ROI encoding approach maintains a comparable level of performance, regardless of the metric used to evaluate the objective quality. Second, the HEVC significantly outperforms the H.264/ AVC compression standard for both codec implementations. In that context, the hevc_nvenc appears to be the best implementation, achieving bitrate demands reductions up to 96% and 91% for ROI methods 0 and 1, respectively. Third, all the tile-based ROI strategies consistently show significant savings compared to full-frame encoding, with bitrate reductions exceeding 74% across all video encoding standards, metrics, and ROI techniques. Across all codecs, ROI methods 1 through 3 show lower BD-Rate savings than the first method (ROI 0), indicating that selecting larger ROI regions results in better quality compression for those areas, but at the cost of slightly higher bitrate requirements.

Table II. Objective BD-RATE Savings (%) Full Frame (QP=22) Vs Tile-Based ROI Techniques (ROI QP=22, Background QP=37), based on PSNR, S-PSNR, WS-PSNR and CPP-PSNR.

PSNR				
	X264	H264 NVENC	X265	HEVC NVENC
ROI METHOD				
0	-85.448	-89.049	-93.91	-94.34
1	-77.383	-81.256	-91.394	-92.166
2	-77.027	-80.344	-91.305	-91.981
3	-74.031	-78.335	-90.91	-91.25
S-PSNR				
0	-81.862	-93.616	-94.293	-96.424
1	-78.865	-87.058	-93.029	-94.431
2	-78.351	-85.590	-92.998	-94.168
3	-75.911	-84.047	-92.425	-93.824
WS-PSNR				
0	-84.819	-91.414	-94.618	-95.329
1	-78.425	-83.913	-92.80	-93.005
2	-77.606	-82.861	-92.604	-92.85
3	-74.866	-81.189	-91.949	-92.321
CPP-PSNR				
0	-83.539	-92.701	-94.486	-95.971
1	-78.652	-85.711	-92.913	-93.843
2	-77.919	-84.489	-93.316	-93.653
3	-75.322	-82.935	-92.173	-93.188

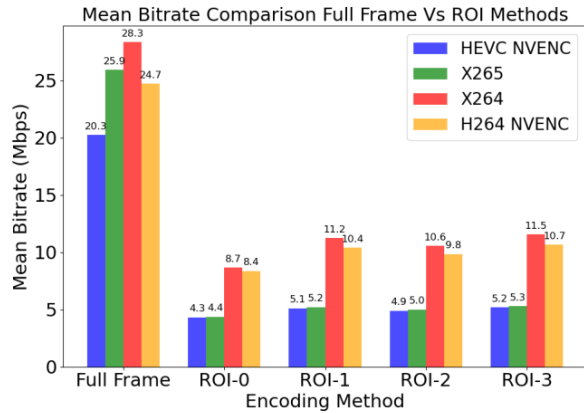


Figure 3. Mean bitrate demands differences between the full-frame encoding method and the four tile-based ROI encoding methods (ROI 0 - ROI 3), for each of the four utilized codecs (hevc_nvenc, x265, h264_nvenc, and x264)

The bar chart in Figure 3 shows the bitrate differences between the full-frame encoding method and the four different tile-based ROI encoding techniques across. The average bitrate was calculated across the 10 videos in our dataset. Among the encoders, hevc_nvenc consistently achieved the lowest bitrates in all scenarios, demonstrating its superior compression efficiency. The x265 is the next best performing codec, outperforming both x264 and h264_nvenc, which is consistent with the existing literature. While x264 showed the lowest overall performance, it still achieved considerable bitrate reductions with the ROI methods.

The results indicate that full-frame encoding requires significantly higher bitrate across all encoders, with hevc_nvenc achieving the highest bitrate (28.3 Mbps), followed by x265 (25.9 Mbps), h264_nvenc (20.3 Mbps) and x264 (24.7 Mbps) all resulting in greater bitrate demands compared to the ROI techniques. This is expected, as full-frame encoding compresses the entire video at a constant, uniform quality level, resulting in higher bitrate demands. In contrast, the tile-based ROI encoding methods result in significantly lower bitrates. Among them, the hevc_nvenc achieves the lowest bitrate overall, closely followed by x265. The bitrate gradually increased, as the ROI level increased, reflecting that the region of interest contains more tiles and thus encoding larger regions at higher quality. For instance, with the ROI-0 method, which has the smallest viewport, the hevc_nvenc codec achieves a mean bitrate of 4.3 Mbps. However, as the viewport size increases with the ROI-3 method, the mean bitrate rises 5.2 Mbps. The differences in bitrate between ROI-0, ROI-1, ROI-2, and ROI-3 are relatively small, suggesting that expanding the high-quality encoded region, as in ROI-3, incurs a relatively small increase in bitrate, which is, however, video content and application specific. Despite this increase, this bitrate remains significantly lower than the bitrates of the full frame encoding, which requires a much higher bitrate of 28.3 Mbps for hevc_nvenc. Similar trends are observed for the rest of the codecs.

IV. DISCUSSION & CONCLUDING REMARKS

This study proposes four tile-based ROI encoding methods, demonstrating their feasibility and effectiveness for 360° video streaming. Using a tiling mechanism to prioritize the quality of the user's viewpoint, these methods help address resource and bandwidth challenges in 360° video delivery.

Across all the examined video encoding standards, HEVC based encoders, particularly hevc_nvenc, consistently outperformed AVC encoders, achieving the lowest bitrates while maintaining high video quality. This compression efficiency aligns with prior literature, confirming HEVC encoders as a suitable choice for applications that require bandwidth optimization.

The comparative analysis of the four ROI methods shows that the slight increase in bitrate from ROI-0 to ROI-3 method highlights an important tradeoff. The ROI-3 method, using a sliding window, encodes a larger ROI at high quality, ensuring a better viewing experience, in contrast to the other methods that encode smaller viewport areas. While the ROI-3 method results in a higher bitrate compared to the lower levels of ROI (0-2), this increase is well within the acceptable limits, with differences remaining significantly below the bitrate demands of full-frame encoding. Linked with perceptual quality assessment, this constitutes ROI-3 as a balanced approach

between bitrate efficiency and user experience. Overall, the tile-based ROI encoding methods achieve substantial bitrate savings compared to full-frame encoding, while maintaining adequately high video quality.

Viewing ROI tile-based encoded videos through an Oculus Meta Quest head movement device allowed us to make initial observations about the visual video quality. We noticed that using the ROI-3 method, the quality within the ROI appeared consistent, with no noticeable differences between full-frame-encoded videos and tile-based ROI videos. However, using a smaller ROI, such as ROI (0-2) techniques, results in pixelation of the background details. This issue demonstrates the ROI method's failure to maintain high-quality rendering for extensive, non-ROI areas, especially when using encoding techniques like x264. Expanding the encoded tile coverage (ROI-3) or using hevc_nvenc improves visibility and reduces pixelation in these areas. Furthermore, in the three motion videos, we observed visual discontinuities at the tile boundaries, which poses a significant problem. While some motion inconsistencies existed in the original videos, the tile-based encoding process amplified these issues, particularly in areas outside the ROI, where motion shifts between tiles became more evident. This indicates that the tiling process may compromise the visual coherence of dynamic scenes if not carefully managed.

In future work, we plan to conduct subjective video quality assessment experiments using head movement devices to validate and complement our objective results. These experiments will also help refine encoding algorithms based on ROI tile-based techniques, addressing challenges such as motion inconsistencies and further enhancing the perceived video quality.

REFERENCES

- [1] O. R. L. Farley, K. Spencer, and L. Baudinet, "Virtual reality in sports coaching, skill acquisition and application to surfing: A review," *jhse*, vol. 15, no. 3, 2019, doi: 10.14198/jhse.2020.153.06.
- [2] M. C. Hsieh and J. J. Lee, "Preliminary Study of VR and AR Applications in Medical and Healthcare Education," *J Nurs Health Stud*, vol. 03, no. 01, 2018, doi: 10.21767/2574-2825.100030.
- [3] J. Pirker and A. Dengel, "The Potential of 360° Virtual Reality Videos and Real VR for Education—A Literature Review," *IEEE Comput. Grap. Appl.*, vol. 41, no. 4, pp. 76–89, Jul. 2021, doi: 10.1109/MCG.2021.3067999.
- [4] "Meta Quest MR, VR Headsets & Accessories." Accessed: Nov. 19, 2024. [Online]. Available: <https://www.meta.com/quest/>
- [5] "Samsung Introduces New Gear VR with Controller, Expanding Gear VR Ecosystem to Make VR Experiences Easier, More Enjoyable." Accessed: Nov. 19, 2024. [Online]. Available: <https://news.samsung.com/global/samsung-introduces-new-gear-vr-with-controller-expanding-gear-vr-ecosystem-to-make-vr-experiences-easier-more-enjoyable>
- [6] "Microsoft HoloLens | Mixed Reality Technology for Business." Accessed: Nov. 19, 2024. [Online]. Available: <https://www.microsoft.com/en-us/hololens>
- [7] Z. Li, Y. Wang, Y. Liu, J. Li, and P. Zhu, "JUST360: Optimizing 360-Degree Video Streaming Systems With Joint Utility," *IEEE Trans. on Broadcast.*, vol. 70, no. 2, pp. 468–481, Jun. 2024, doi: 10.1109/TBC.2024.3374066.
- [8] M. Mahmoud *et al.*, *Versatile Video Coding Performance Evaluation for Tiled 360° Videos*. EUROPEAN WIRELESS, 2023.
- [9] A. Dharmasiri, C. Kattadige, V. Zhang, and K. Thilakarathna, "Viewport-Aware Dynamic 360° Video Segment Categorization," May 13, 2021, *arXiv:2105.01701*. Accessed: Nov. 19, 2024. [Online]. Available: <http://arxiv.org/abs/2105.01701>
- [10] C. Kattadige and K. Thilakarathna, "VAD360: Viewport Aware Dynamic 360-Degree Video Frame Tiling," May 24, 2021, *arXiv:2105.11563*. Accessed: Nov. 19, 2024. [Online]. Available: <http://arxiv.org/abs/2105.11563>
- [11] M. Mahmoud *et al.*, "A Survey on Optimizing Mobile Delivery of 360° Videos: Edge Caching and Multicasting," *IEEE Access*, vol. 11, pp. 68925–68942, 2023, doi: 10.1109/ACCESS.2023.3292335.
- [12] F. Qian, B. Han, Q. Xiao, and V. Gopalakrishnan, "Flare: Practical Viewport-Adaptive 360-Degree Video Streaming for Mobile Devices," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, New Delhi India: ACM, Oct. 2018, pp. 99–114. doi: 10.1145/3241539.3241565.
- [13] M. Mahmoud *et al.*, "Optimized Tile Quality Selection in Multi-user 360° Video Streaming," *IEEE Open J. Commun. Soc.*, pp. 1–1, 2024, doi: 10.1109/OJCOMS.2024.3498334.
- [14] M. Xiao, C. Zhou, Y. Liu, and S. Chen, "OpTile: Toward Optimal Tiling in 360-degree Video Streaming," in *Proceedings of the 25th ACM international conference on Multimedia*, Mountain View California USA: ACM, Oct. 2017, pp. 708–716. doi: 10.1145/3123266.3123339.
- [15] C. Li, M. Xu, X. Du, and Z. Wang, "Bridge the Gap Between VQA and Human Behavior on Omnidirectional Video: A Large-Scale Dataset and a Deep Learning Model," in *Proceedings of the 26th ACM international conference on Multimedia*, Seoul Republic of Korea: ACM, Oct. 2018, pp. 932–940. doi: 10.1145/3240508.3240581.
- [16] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003, doi: 10.1109/TCSVT.2003.815165.
- [17] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012, doi: 10.1109/TCSVT.2012.2221191.
- [18] "Encode/H.264 - FFmpeg." Accessed: Aug. 08, 2024. [Online]. Available: <https://trac.ffmpeg.org/wiki/Encode/H.264>
- [19] "Encode/H.265 - FFmpeg." Accessed: Nov. 21, 2024. [Online]. Available: <https://trac.ffmpeg.org/wiki/Encode/H.265>
- [20] "Using FFmpeg with NVIDIA GPU Hardware Acceleration," NVIDIA Docs. Accessed: Nov. 21, 2024. [Online]. Available: <https://docs.nvidia.com/video-technologies/video-codec-sdk/11.1/ffmpeg-with-nvidia-gpu/index.html>
- [21] *Samsung/360tools*. (Nov. 19, 2024). C. Samsung. Accessed: Nov. 21, 2024. [Online]. Available: <https://github.com/Samsung/360tools>
- [22] I. Valiandi, A. Panayides, E. Kyriacou, C. Pattichis, and M. Pattichis, "A Comparative Performance Assessment of Different Video Codecs," 2023, pp. 265–275. doi: 10.1007/978-3-031-44240-7_26.
- [23] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards—Including High Efficiency Video Coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1669–1684, Dec. 2012, doi: 10.1109/TCSVT.2012.2221192.
- [24] M. Yu, H. Lakshman, and B. Girod, "A Framework to Evaluate Omnidirectional Video Coding Schemes," in *2015 IEEE International Symposium on Mixed and Augmented Reality*, Fukuoka, Japan: IEEE, Sep. 2015, pp. 31–36. doi: 10.1109/ISMAR.2015.12.
- [25] V. Zakharchenko, K. P. Choi, E. Alshina, and J. H. Park, "Omnidirectional Video Quality Metrics and Evaluation Process," in *2017 Data Compression Conference (DCC)*, Snowbird, UT, USA: IEEE, Apr. 2017, pp. 472–472. doi: 10.1109/DCC.2017.90.
- [26] V. Zakharchenko, K. P. Choi, and J. H. Park, "Quality metric for spherical panoramic video," presented at the SPIE Optical Engineering + Applications, K. M. Iftikharuddin, A. A. S. Awwal, M. García Vázquez, A. Márquez, and M. A. Matin, Eds., San Diego, California, United States, Sep. 2016, p. 99700C. doi: 10.1117/12.2235885.
- [27] Y. Sun, A. Lu, and L. Yu, "Weighted-to-Spherically-Uniform Quality Evaluation for Omnidirectional Video," *IEEE Signal Process. Lett.*, pp. 1–1, 2017, doi: 10.1109/LSP.2017.2720693.