# A Coding-Efficiency-Aware Fast Heuristic for VVC Intra-Frame Prediction Targeting 360° Videos

Rogério Rosado, Otávio Santos
Franklin Oliveira, Lucas Silva
{rcassanta,otavio.fsantos}
{fsoliveira,lsdsilva}@inf.ufpel.edu.br
Video Technology Research Group
Federal University of Pelotas, Brazil

Gustavo Sanchez
gustavo.sanchez@ifsertao-pe.edu.br
Federal Institute of Sertão
Pernambucano
Salgueiro/PE, Brazil

Iago Storch, Daniel Palomino
Luciano Agostini
{icstorch,dpalomino,agostini}
@inf.ufpel.edu.br
Video Technology Research Group
Federal University of Pelotas, Brazil

## ABSTRACT

360° videos have gained prominence among immersive media, providing users with an impressive immersive scene experience. But the computational effort necessary to encode these videos is very high since the 360° videos demand more data to be represented. This paper presents a heuristic-based decision solution for the intra-frame prediction in Versatile Video Coding (VVC) focused on 360° videos. The main goal is to reduce the computational effort of the intra-frame prediction process while ensuring minimal impact on coding efficiency. This is achieved exploring the distortions introduced by the equirectangular projection used to encode 360° videos. The results of the proposed heuristic present time savings of 8.34% with a minor coding efficiency loss of only 0.16% BD-BR. This loss is the smallest among related works, allowing the proposed solution to achieve the best balance between time savings and coding efficiency losses among works in the literature.

## KEYWORDS

360° videos, VVC, fast intra-frame prediction, coding-efficiency-aware.

## 1 INTRODUCTION

Nowadays, digital videos are increasingly popular and used in various applications like entertainment, video calls, and e-learning, among others. Due to the constant technological advancements in video acquisition, display, and processing, the industry is investing in higher resolutions and interactive and immersive approaches for digital videos, including 360° videos, also called omnidirectional videos. In recent years, there has been rapid development in Virtual Reality (VR) and immersion technologies, and 360° videos play an important role in this process [14].

One of the biggest challenges of this type of content is the high bandwidth required for high quality delivery. The 360° video content represents a sphere that covers the entire 360°×180° field of view. In other words, the 360° video envelops the viewer perfectly and occupies their entire vision, which differs from traditional two-dimensional video that covers only a limited plane [3].

Currently, modern video codecs search for new solutions to obtain compression gains. Among these codecs we highlight VVC (Versatile Video Coding) [13] [3], finalized in July 2020, as the most recent video coding standard. VVC introduces important new tools in all stages of the encoding process, including new partitioning types, which will be explored in this work. These new encoding tools makes the computational effort of VVC much higher than other standards. Therefore, it is important to develop solutions to reduce this computational cost.

The capture of 360° videos uses a 360° camera with several conventional 2D cameras to capture the video from a wide field of view. After the capture, the image is sewed to obtain a spherical representation of the scene. The 360° video is projected into a 2D plane using a projection format to be encoded. The most used format, and the focus of this work, is the equirectangular projection (ERP), also called cylindrical projection [15]. One example of ERP is presented in Figure 2. ERP is a simple cartographic projection that maps the meridians in vertical lines with constant spacing and latitude circles with straight horizontal lines with constant spacing. The ERP format includes some distortions in the images, which tend to increase with the distance of the image center (equator). This characteristic will be explored in this work to reduce the computational effort required by VVC intra-frame prediction when encoding 360° videos. After the projection, the video is encoded with a codec for 2D video, such as VVC [13] [3].

There are some works in the literature focusing on reducing the computational effort of intra-frame prediction in VVC by exploring the characteristics of 360° videos. Some of these works use machine learning approaches, such as [4], [12] and [18] to achieve more aggressive results in terms of computational effort reduction. Other works, such as [17] and [8], propose heuristic-based solutions to reduce computational costs without machine learning. Some works also focus on the HEVC encoder, such as [1] and [10]. In general, most works in the literature focus more on reducing the computational effort with less focus on preserving coding efficiency. In this context, while it is necessary to develop solutions that reduce the computational effort of the intra-frame prediction in VVC, they also need to keep the coding efficiency loss as minimal as possible.

This work presents a heuristic-based decision for the intra-frame prediction in VVC targeting 360° videos. The proposed solution focus on reducing the computational effort while preserving coding efficiency with minimal loss. The approach of the proposed solution lies in the hypothesis that in the VVC partition process [7], horizontal rectangular block formats are more relevant than vertical rectangular blocks in the upper and lower regions of 360° videos when the equirectangular projection (ERP) [15] is used.

In: Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia'2024). Juiz de Fora, Brazil. Porto Alegre: Brazilian Computer Society, 2024.

ISSN 2966-2753

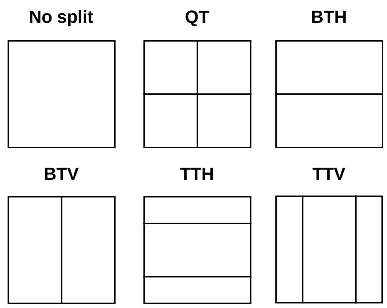


**Figure 1: Partitioning types available in QTMT**


**Figure 2: Division of the frame into bands.**

## 2 INTRA-FRAME PREDICTION IN VVC

VVC standard divides each input frame into Coding Tree Blocks (CTBs) with luminance and chrominance samples from the region of the frame. Each CTB covers a square region of up to 128x128 luminance samples and the respective chrominance samples. Each CTB has three Coding Tree Units (CTUs), one for each color matrix: luminance (Y), blue (Cb) and red (Cr) chrominance [7].

Each CTU can be recursively partitioned into smaller blocks defined as Coding Units (CUs). The division process is based on a coding tree structure known as Quadtree with a nested multi-type tree (QTMT). This structure is formed by a Quaternary Tree (QT), as in other standards, where each leaf node of the QT is the root of a Multi Type Tree (MTT), which allows binary or ternary divisions. Therefore, the six types of operations shown in Figure 1 are allowed. A block can be encoded without divisions (No Split in Figure 1), or it can be divided with the QT, BTH, BTV, TTH, or TTV partitions[7].

The CTU is first partitioned recursively with a QT structure. Subsequently, each leaf node of the QT tree can be partitioned further with a recursive MTT structure using binary or ternary divisions. However, when an MTT division is performed, a subsequent QT division is no longer allowed. The CU sizes can range from $4 \times 4$ to $128 \times 128$ samples (maximum CTU size), with 28 block sizes allowed. The leaf nodes of the QTMT represent the CUs, which are the units that are effectively encoded [9].

VVC intra-frame coding takes advantage of the partitioning structure and includes new coding tools to enhance coding efficiency compared to previous standards. A new tool called dual tree is introduced for block partitioning, which allows separate partitioning trees for luminance and chrominance. The number of angular prediction modes is increased to 67, up from 35 in HEVC. This set consists of modes DC, planar, and 65 directional modes. Furthermore, the Wide Angle Intra-frame Prediction (WAIP) tool was introduced in VVC, adapting angular modes for non-square blocks. The Multiple Reference Lines (MRL) tool was also introduced, which allows the encoder to choose between three reference lines in the prediction instead of just one, as in previous standards. Another new feature was the Intra Subpartion (ISP) tool, which allows an additional division into two or four sub-blocks for blocks processed by the intra-frame prediction. Finally, the Matrix-Based Intra-frame Prediction (MIP) tool, designed using artificial intelligence, was introduced. This tool can replace conventional prediction modes by multiplying the reference samples with predefined matrices [5][7].

## 3 ANALYSIS ON VVC INTRA-FRAME PREDICTION FOR 360º VIDEOS

As previously discussed, ERP projection distorts the image, especially in upper and lower regions. This section presents experiments to investigate the effects of these distortions in VVC intra-frame prediction, considering different regions of the image. In particular, these experiments intend to evaluate the relevance of rectangular blocks (partitions BTH, BTV, TTH e TTV in Figure 1) in these regions where the ERP projection applies more distortion. Initially, each video frame was divided into three bands, as proposed in [10]: upper band, middle band, and lower band. This division is exemplified in Figure 2. The middle band has 50% of the frame samples, while the upper and lower bands each have 25% of the samples.

The first experiment sought to verify the percentage of blocks of different formats used to encode each band of video frames, considering the intra-frame prediction. For this, three categories were considered: (i) square blocks (No Split partition as in Figure 1), (ii) horizontal blocks (BTH and TTH partitions as in Figure 1) and (iii) vertical blocks (BTV and TTV partitions as in Figure 1).

This experiment used the reference software for the VVC standard, the VVC Test Model or VTM [11], version 19.0. The All Intra configuration was used (all frames are encoded with intra-frame), and the Quantization Parameter (QP) was set to 22. The video sequence was SkatboardInLot, which has 300 frames in an $8192 \times 4096$ resolution, with 10 bits per sample [6]. VTM was modified to count the blocks chosen by intra-frame prediction in each frame band.

The results are summarized in the graph shown in Figure 3. In all cases, square blocks are heavily used. In the middle band, square, horizontal, and vertical blocks are relatively balanced. On the other hand, there is a clear imbalance in the upper and lower regions. The main hypothesis for the heuristic-based decision developed in this work is that, at the intra-frame processing, horizontal rectangular blocks are more frequently used than vertical blocks in the upper and lower bands of the frames. Experimental data demonstrate that this hypothesis is true. On the upper band, horizontal blocks are selected 11 times more often than vertical blocks. This difference is smaller in the lower band, where horizontal blocks are chosen about 8.5 times more often than vertical blocks. Yet, this is still a substantial imbalance when compared to the middle band behavior.

A second set of experiments was conducted to verify the impact on coding efficiency if vertical and horizontal partition types are disabled in the upper and lower bands. A more complete experimental

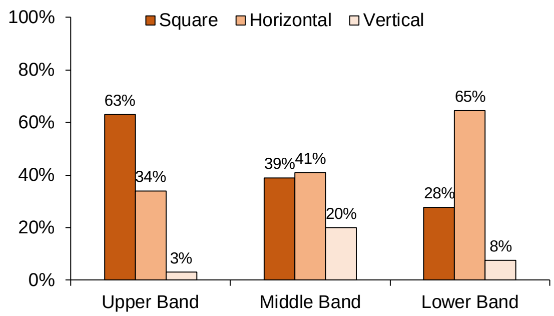


**Figure 3: Use of squared, horizontal and vertical blocks per band in the SkatboardInLot video (QP 22)**

set was employed in this case, using the eight video sequences defined by the VVC Common Test Conditions (CTC) for experiments with 360° videos [6]. The sequences were ChairliftRide, Gaslamp, Harbor, KiteFlite, SkateboardInLot, SkateboardTrick, Train e Trolley. Additionally, four QPs were used: 22, 27, 32 e 37, as defined by the CTCs. Again, the All Intra configuration was chosen.

Two metrics were used for this evaluation: Bjontegaard Delta Bitrate (BD-BR) and Bjontegaard Delta Weighted-to-Spherically Uniform PSNR (BD-WSPSNR). These are the most commonly used metrics to evaluate the coding efficiency of 360° videos. BD-BR is a metric that compares two encoders, indicating, for the same objective quality, the increase or decrease in bitrate needed to represent the video [2]. For 360° videos, objective quality is measured by the WS-PSNR metric [10]. On the other hand, BD-WSPSNR considers the impact on objective quality for the same bitrate. Again, WS-PSNR is used in the context of 360° videos. WSPSNR is a weighted version of PSNR that considers the projection distortion and better represents the objective quality in projected 360° videos [16].

Table 1 presents the detailed results of this experiment, considering the average values for the four QPs across all evaluated video sequences. The second and third columns show the experiment's results of disabling the horizontal partition types (i.e., BTH and TTH), while the fourth and fifth columns show the results when vertical partitions (i.e., BTV and TTV) are disabled

The results presented in Table 1 reinforce the validity of the hypothesis that vertical partitions are of little relevance in the upper and lower bands of the images. Removing vertical partitions from these regions caused an average impact of only 0.16% on BD-BR and only -0.71 dB on BD-WSPSNR. On the other hand, removing horizontal partitions caused much more significant impacts, with a 0.77% impact on BD-BR and -4.07 dB on BD-WSPSNR. This implies a 4.8 times greater impact on BD-BR and a 5.7 times greater impact on BD-WSPSNR when horizontal partitions are removed. These results boosted the use of the heuristic proposed in this work, which is better discussed in the next section of this paper.

Another important conclusion from Table 1 results is that it is possible to observe a variation in losses depending on video content, which is expected behavior. Videos with more homogeneous textures in the lower and/or upper bands, such as a sky, tend to have smaller losses in both experiments, as these regions are typically encoded with large square blocks. On the other hand, if there are

**Table 1: Coding efficiency when horizontal and vertical partition types are disabled at the upper and lower bands.**

| Video | No Horizontal | | No Vertical | |
|---|---|---|---|---|
| | BD-BR (%) | BD-WS PSNR | BD-BR (%) | BD-WS PSNR |
| ChairliftRide | 0.88 | -0.032 dB | 0.19 | -0.006 dB |
| Gaslamp | 0.96 | -0.044 dB | 0.37 | -0,011 dB |
| Harbor | 0.98 | -0,047 dB | 0.35 | -0,016 dB |
| KiteFile | 1.61 | -0,103 dB | 0.10 | -0,006 dB |
| SkateboardInlot | 0.26 | -0,011 dB | 0.08 | -0,003 dB |
| SkateboardTrick | 0.24 | -0,011 dB | 0.04 | -0,005 dB |
| Train | 0.90 | -0,018 dB | 0.06 | -0,002 dB |
| **Average** | **0.77** | **-0,040 dB** | **0.16** | **-0,007 dB** |

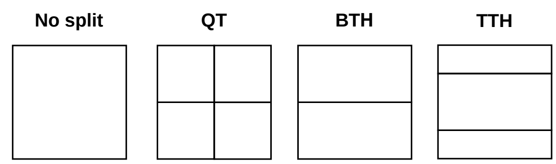


**Figure 4: Partitioning types available in QTMTp.**

more elaborate textures on the lower and upper bands, the exclusion of horizontal partitions tends to have a greater impact since there is horizontal stretching in these regions, making horizontal blocks more relevant for encoding these areas.

## 4 PROPOSED HEURISTIC SOLUTION

Considering the experiments presented in the previous section, it was demonstrated that the initial hypothesis of this work is true. Therefore, a simple heuristic based on this hypothesis was proposed: disable vertical partitions in the upper and lower bands of the frame. Then, these frame regions will use a simplified version of the QTMT, called QTMTp, in this work, where "p" indicates its use in the polar regions, namely the upper and lower bands of the frame. In fact, the simplification occurs at the MTT, since binary vertical (BTV) and ternary vertical (TTV) partitioning are not available at the QTMTp tree. Then, the partitioning types available at QTMTp are shown in Figure 4. In the middle band of the frames, the QTMT is used and all partitioning types are available.

The simplified diagram of the proposed heuristic is shown in Figure 5. The orange steps are added to the regular VVC intra-frame prediction process. The position of each processed CTU is evaluated to define whether the partition will be performed with the VVC original QTMT (for the middle band) or with the QTMTp (for the upper and lower bands).

Since the upper and lower bands together contain 50% of the frame area, this simplification will affect half the area of all frames. This solution was implemented in the reference software and its results are presented in the next section.

## 5 RESULTS

The proposed heuristic-based mode decision solution presented in the previous section was implemented in VTM, version 19.0. It was

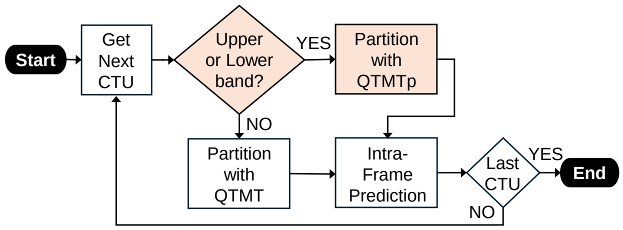


**Figure 5: Diagram illustrating the operation of the proposed heuristic.**

**Table 2: Coding efficiency and execution time results of the proposed heuristic.**

| Video | BD-BR (%) | TS (%) |
|---|---|---|
| ChairliftRide | 0.19 | 9.6 |
| Gaslamp | 0.37 | 7.7 |
| Harbor | 0.35 | 9.9 |
| KiteFlite | 0.10 | 10.6 |
| SkateboardInLot | 0.08 | 8.9 |
| SkateborardTrick | 0.11 | 3.1 |
| Train | 0.04 | 6.5 |
| Trolley | 0.06 | 10.5 |
| **Average** | **0.16** | **8.34** |

evaluated using the same experimental method presented in Section 4, but now including the complete set of video sequences and QPs defined in the CTC recommendations for 360° videos, which is the reference in the area for experiments with VVC processing 360° videos. Once again, the All Intra configuration was used. The VTM was also modified to insert coding time counters to allow the evaluation of the proposed method's encoding time.

The results are presented in Table 2. In this table, the BD-BR results from Table 1 were retained and the execution time savings (TS) results were included. These results show the percentage reduction in execution time with the proposed heuristic compared to the original implementation of VTM. It is worth noting that this gain considers the execution time of the entire encoder and not just intra-frame prediction execution time.

Observing Table 2, it is possible to conclude that the proposed heuristic presented gains in encoding time in all evaluated videos. These gains vary depending on video content, ranging from 3.1% to 10.6%, with an average of 8.34%. These results were considered very good, as a significant reduction in the VTM encoder execution time was achieved with a very low impact on coding efficiency, which was the main goal of this work.

Coding efficiency losses ranged from 0.04% to 0.37%, with an average of 0.16% BD-BR. Therefore, the relationship between execution time reduction and coding efficiency loss is highly favorable, demonstrating the relevance of the obtained results. The reasons for these results are primarily twofold: (i) the significance of the initial hypothesis, experimentally validated in this work, and (ii) the simplicity of the proposed method, which has a negligible time overhead to be implemented within the encoder.

**Table 3: Comparison with related works.**

| Work | BD-BR (%) | TS (%) | TS/BD-BR |
|---|---|---|---|
| Zhang [17] | 0.66 | 32.13 | 48.68 |
| Liu [8] | 1.40 | 38.73 | 27.66 |
| Filipe [4] | 1.46 | 57.45 | 39.35 |
| Zhewen [12] | 1.78 | 72.17 | 40.54 |
| Shu [18] | 1.96 | 60.40 | 30.82 |
| **This work** | **0.16** | **8.34** | **52.13** |

The comparison with related works is presented in Table 3, where the coding efficiency losses (BD-BR) and the time savings (TS) of the different solutions are presented. It is important to note that the works used different versions of VTM and also different video sequences in their experiments, making the comparisons not entirely fair. Nevertheless, by examining the results in Table 3, it is possible to conclude that the solution presented in this work achieved the lowest coding efficiency loss, which was the main objective of this work. The other works present coding efficiency losses between 4.1 and 14.7 times greater than ours in terms of BD-BR. On the other hand, more aggressive strategies for reducing computational costs in related works allowed them to achieve time saving results between 3.8 and 8.6 larger than the proposed solution.

Table 3 also presents the relationship between TS and BD-BR. This relationship evaluates both metrics together. In this regard, the method proposed in this work presented the best results, as it is the one that achieves important gains in reducing computational effort with minimal impacts on coding efficiency.

# 6 CONCLUSION

This work presented a heuristic-based solution to reduce the computational effort of intra-frame prediction in the VVC standard when processing 360° videos. This heuristic had the objective of posing negligible coding efficiency penalties. The proposed heuristic defines a specific partitioning tree for upper and lower bands of the video frames, called QTMTp, where vertical partition types are not allowed. The heuristic was implemented in VTM and the results showed average time savings of 8.34% with small coding efficiency losses of only 0.16% BD-BR. When compared with works in the literature, the proposed heuristic was the one that presented the lowest losses in coding efficiency and the one that presented the best relationship between time savings and coding efficiency. In future works, other heuristics will be investigated considering the directionality of the angular prediction modes and the intra-subpartition tool, both from VVC intra-frame prediction.

# ACKNOWLEDGMENTS

The authors of this work would like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) Finance Code 001, CNPq, FAPERGS, FACEPE and IFFar for funding this research.

## REFERENCES


[1] Bernardo Beling, Iago Storch, Luciano Agostini, Bruno Zatt, Sergio Bampi, and Daniel Palomino. 2020. ERP-Based CTU Splitting Early Termination for Intra Prediction of 360 videos. In *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. 359–362. https://doi.org/10.1109/VCIP49819.2020.9301879

[2] Gisle Bjontegaard. 2001. Calculation of average PSNR differences between RD-curves. Technical Report VCEG-M33, ITU-TSG16/Q6, Austin, Texas, USA.

[3] Benjamin Bross, Ye-Kui Wang, Yan Ye, Shan Liu, Jianle Chen, Gary J. Sullivan, and Jens-Rainer Ohm. 2021. Overview of the Versatile Video Coding (VVC) Standard and its Applications. *IEEE Transactions on Circuits and Systems for Video Technology* 31, 10 (2021), 3736–3764. https://doi.org/10.1109/TCSVT.2021.3101953

[4] Jose Filipe, João Carreira, Luis Tavora, Sergio de Faria, Antonio Navarro, and Pedro Assuncao. 2021. Tree-Based Ensemble Methods for Complexity Reduction of VVC Intra Coding. In *2021 Telecoms Conference (ConfTELE)*. 1–6. https://doi.org/10.1109/ConfTELE50222.2021.9435476

[5] Wassim Hamidouche, Thibaud Biatek, Mohsen Abdoli, Edouard François, Fernando Pescador, Miloš Radosavljević, Daniel Menard, and Mickael Raulet. 2022. Versatile Video Coding Standard: A Review From Coding Tools to Consumers Deployment. *IEEE Consumer Electronics Magazine* 11, 5 (2022), 10–24. https://doi.org/10.1109/MCE.2022.3144545

[6] Philippe Hanhart, Jill Boyce, Kiho Choi, and Jian-Liang Lin. 2018. JVET common test conditions and evaluation procedures for 360° video. JVET-L1012-v1.

[7] Yu-Wen Huang, Jicheng An, Han Huang, Xiang Li, Shih-Ta Hsiang, Kai Zhang, Han Gao, Jackie Ma, and Olena Chubach. 2021. Block Partitioning Structure in the VVC Standard. *IEEE Transactions on Circuits and Systems for Video Technology* 31, 10 (2021), 3818–3833. https://doi.org/10.1109/TCSVT.2021.3088134

[8] Zhi Liu, Cai Xu, Mengmeng Zhang, and Xiaohan Guan. 2019. Fast Intra Prediction Algorithm for Virtual Reality 360 Degree Video Based on Improved RMD. In *2019 Data Compression Conference (DCC)*. 593–593. https://doi.org/10.1109/DCC.2019.00105

[9] Mário Saldanha, Gustavo Sanchez, César Marcon, and Luciano Agostini. 2021. Performance analysis of VVC intra coding. *Journal of Visual Communication and Image Representation* 79 (2021), 103202. https://doi.org/10.1016/j.jvcir.2021.103202

[10] Iago Storch, Bruno Zatt, Luciano Agostini, Guilherme Correa, Luis A. da Silva Cruz, and Daniel Palomino. 2020. Spatially Adaptive Intra Mode Pre-Selection for ERP 360 Video Coding. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2178–2182. https://doi.org/10.1109/ICASSP40776.2020.9053374

[11] Karsten Suehring. 2022. VTM-19.0. https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/releases/VTM-19.0 Accessed: 2024-06-28.

[12] Zhewen Sun, Li Yu, and Wei Peng. 2023. QTMT-LNN: A fast intra CU partition using lightweight neural network for 360-degree video coding on VVC. *IET Image Processing* 17, 2 (2023), 597–612. https://doi.org/10.1049/ipr2.12658

[13] International Telecommunication Union. 2023. Recommendation ITU-T H.266 (V3) Versatile video coding. https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-H.266-202309-I!!PDF-E&type=items. Accessed: 2024-06-28.

[14] Mai Xu, Chen Li, Shanyi Zhang, and Patrick Le Callet. 2020. State-of-the-Art in 360° Video/Image Processing: Perception, Assessment and Compression. *IEEE Journal of Selected Topics in Signal Processing* 14, 1 (2020), 5–26. https://doi.org/10.1109/JSTSP.2020.2966864

[15] Yan Ye, Jill M. Boyce, and Philippe Hanhart. 2020. Omnidirectional 360° Video Coding Technology in Responses to the Joint Call for Proposals on Video Compression With Capability Beyond HEVC. *IEEE Transactions on Circuits and Systems for Video Technology* 30, 5 (2020), 1241–1252. https://doi.org/10.1109/TCSVT.2019.2953827

[16] Matt Yu, Haricharan Lakshman, and Bernd Girod. 2015. A Framework to Evaluate Omnidirectional Video Coding Schemes. In *2015 IEEE International Symposium on Mixed and Augmented Reality*. 31–36. https://doi.org/10.1109/ISMAR.2015.12

[17] Mengmeng Zhang, Yan Hou, and Zhi Liu. 2023. An early CU partition mode decision algorithm in VVC based on variogram for virtual reality 360 degree videos. *EURASIP Journal on Image and Video Processing* 2023 (2023), 9. https://doi.org/10.1186/s13640-023-00605-0

[18] Shu Zheng-Jie, Peng Zong-Ju, Jiang Gang-Yi, Chen Fen, and Yuan Bo-Sen. 2023. Fast intra partition and mode prediction for equirectangular projection 360-degree video coding. *IET Image Processing* 17, 2 (2023), 558–569. https://doi.org/10.1049/ipr2.12655