COMPARISON OF BOUNDARY ARTIFACT REMOVAL METHODS IN CODING OF GENERALIZED CUBEMAP PROJECTION USING VVC

Kianoush Jafari, Alireza Aminlou, Miska M. Hannuksela

Nokia Technologies, Tampere, Finland

ABSTRACT

Virtual reality applications use 360-degree videos and head mount displays with stereoscopic capabilities to provide full immersion experience. Among existing projection format, Cubemap projection provides and improved compression performance, but similar to other projects, it suffers from visual artifacts in the rendered viewport because of discontinuity of the content in different regions. In this work, we investigated the effect of different methods in Versatile Video Coding standard and other pre- and post-processing algorithms for removing boundary artifacts by introducing a new objective quality metric for systematic comparison. We further improve the current methods by aligning the GCMP's face boundaries to coding unit boundaries. The observation is that the combination of these method with existing methods in VVC offers the best result.

Index Terms— 360-degree video coding, Virtual reality, Versatile Video Coding (VVC), Generalized Cubemap projection (GCMP)

1. INTRODUCTION

With the emergence of VR as a new rising star of immersive medias and increasing popularity of head mounted display (HMD) the demand for streaming and encoding high-quality omnidirectional video content has been on record high in the last few years. The challenge in this area mainly comes from the fact that the current standards are designed from ground-up to work with 2-D frames not spherical videos. Hence, the 360-degree content must be projected on the 2-D images before any encoding could be done. In this regard, several different formats and frame packing schemes were designed for sampling spherical content and projecting it on the 2-D frame.

Equirectangular projection (ERP) and Cubemap Projection (CMP) are the two of the most important omnidirectional formats. ERP is more well-known and traditional format, while CMP is a more recently developed one. ERP uses a single face to map the sphere on a 2-D plane, while CMP utilizes six to achieve the same. In addition, CMP has been shown to give 5% to 10% more bitrate reduction compare to ERP [1] and have the advantage of being extensively adapted by industry [2].

In CMP, the adjacent faces might have discontinued contents depending on frame packing scheme. This usually would lead to some visible artifact on the boundary of the faces after decoding and viewport rendering. The artifact is generated during encoding either by applying in-loop filtering and motion compensation on connected areas with irrelevant contents (discontinuity boundary) or by different coding on disconnected areas with spatially relevant content. The latter is called seam artifact, while the former is being referred to as leaking artifact. In VVC standard [3], there are encoding tools and methods, such as virtual boundary (VB), subpicture (SP), and face boundary padding, to deal with the boundary artifact problem. Considering this, several subjective studies have been done to evaluate the performance of these methods [4], but so far, no quality metric has been introduced to analysis and compare these methods objectively.

In this paper, we introduce a new boundary objective metric to systematically analyze and compare the performance of different methods, in overall and in face boundary regions. The goal is to use this metric to perform a comprehensive study on a wide spectrum of methods and investigate their effectiveness on removing CMP boundary artifacts. In this regard, we first analyze existing tools which are available in VVC standard and then proceed with proposing new ideas to improve the current performance.

The organization of rest of the paper is as follows: In the section 2, we give a brief background on CMP and Generalized CMP (GCMP) formats. In section 3, we introduce our methods for improving GCMP and our boundary objective metric. In section 4, we present experimental results, and we conclude the paper in section 5.

2. BACKGROUND

2.1. Cubemap projection (CMP)

The CMP places the sphere within a cube and project it on each face by intersecting the rays projected outward from the center of the sphere to the faces of the cube. After the projection, the cube's faces are unfolded and packed to form a 2-D frame. The arrangement of the faces, however, can be done in multiple different ways with 3x2 packing being the most popular one since it produces a more compact and bandwidth-efficient frame. The faces can also be rotated to ensure the maximum possible continuity of content within the frame.

2.2. Generalized CMP (GCMP)

Even though rotating faces would make some improvements, the amount of discontinuity remains high within the CMP frame. Moreover, the CMP itself is not entirely out of distortion, specially at face boundaries. In order to deal with these issues, a range of additional formats, such as GCMP, are introduced which provides non-linear mapping functions as well as different types of paddings as a means of distortion and continuity improvement respectively.

GCMP is a more versatile version of the CMP. While supporting CMP as the main projection format, it has several additional modes which provides different mapping functions as well as various types and schemes of padding. Most recently, blending feature [5] has been added to GCMP to alleviate boundary artifacts at discontinuity areas. The basic idea, as show in Figure 1 for example, is to add spherical padding to some of the discontinue areas before encoding, and then blend those areas, which now have been coded twice in different borders of the image, with weighted averaging.

3. PROPOSED METHODOLOGY

There have been several activities showing that different coding tools such as subpicture, virtual boundary, and padding samples in GCMP visually reduce the boundary artifacts, but there has not been any objective method to quantify this performance improvement. The goal of this research it to perform a comprehensive and systematic analysis on Rate-Distortion (RD) performance of these tools. To this end, we perform several experiments for different methods and compare their outcome using our own developed objective quality metric which will be introduced in this section. Then we propose an improvement on top of the existing padded GCMP.

3.1. Proposed Quality Metric

There exist several common metrics to assess the quality of entire 360-degree content such as S-PSNR, WS-PSNR, and CPP-PSNR [6]. For this study, however, we have proposed a new objective metric, which emphasizes quality computation on a narrow area of CMP face boundaries. The general diagram of our metric has been depicted in Figure 2 for each CMP face. According to this figure, quality (i.e. PSNR) is computed at two different layers for every face boundary. The first is very narrow rectangular regions (green area) with the same length as face size and width of four samples, covering the first four samples from the border toward the center. The other layer covers central area (orange area) separately for each face.

In this study, we need to compare RD performance of different methods having different CMP face resolution. Hence, to have a fair comparison, we use end-to-end quality metric suggested in [8]. In this method, the original input content with higher resolution is converted to a coding content having lower resolution for coding. After decoding and applying blending if needed, the decoded content is

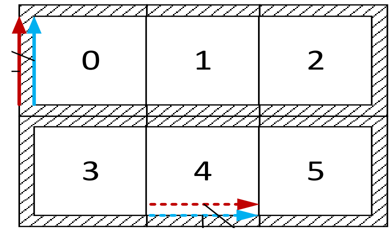

Figure 1: Padded areas in CMP-based formats [5]

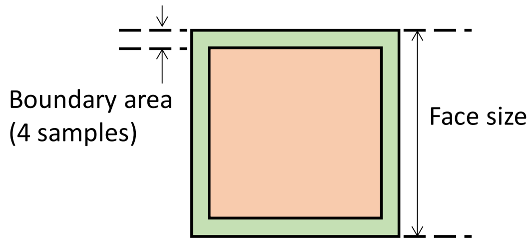

Figure 2: Regions for boundary PSNR in each CMP face


Figure 3: Location of subpicture (left) and VB (right)

converted back to the original resolution. The PSNR values are calculated between the unsampled and the original contents, similar to end-to-end quality metrics.

3.2. Coding Methods

Our first approach is to use VVC's built-in tools, which are available in related software packages such as VTM-11.0. Subpicture is a VVC feature which performs picture partitioning at highest level [9]. Subsequently, each partition is coded independently and transmitted through the network. In the case, as shown Figure 3-left, we divide the 3x2 CMP frame into two equally sized horizontal subpictures with their shared border being the discontinuity boundary. We also enabled subpicture boundary padding feature.

The next VVC tool to examine is VB which defines an imaginary line across the picture which disables in-loop filtering across that line [10]. In our application, as shown in Figure 3-right, we define VB right across the discontinuity boundary to prevent leaking between two rows.

Using subpicture or virtual boundary features, reduced boundary artifact, but there is still visible artifact in the rendered viewport in the in the discontinuity area. To address this issue, we configure GCMP's settings to have spherical padding areas added similar to Figure 1. The padded areas are later blended after decoding to produce smoother face boundaries with significantly less artifact. The downside of this method is the increased bitrate due to extra padding around CMP frames.

In order to reduce the bitrate in the GCMP, which has been increased due to extra padding area, as well as addressing both leaking and seam artifact in a single method, we propose a new approach called aligning, in which a smaller but CTU-aligned frame size can be combined with subpicture (or VB) to produce a better result. In aligning, the new frame height has the same size as initial frame height before padding, while the frame width is slightly reduced.
To create and aligned GMP (AGCMP) frame, we resize the initial face size in such a way that once the padded areas were added, frame's height would stay unchanged. This would allow the frame height stay CTU-aligned during the process. This is especially critical for combining AGCMP with subpicture since the later has to be placed on CTU aligned discontinuity boundary, which does not exist in the case of simple GCMP.

4. EXPERIMENTAL RESULTS

The simulations are performed using 300 frames original uncompressed 360-degree video test sequences, listed in Table 1. As described in [11], the 6K and 8K 360-degree video test sequences, defined in [12], were first converted from ERP to CMP format with a face size of 1536×1536 and 2048×2048, respectively, using JVET 360Lib-12.0 software [13]. The subpicture sequences were coded with VTM 11.0 [14] with face size of 1280×1280, using random access (RA) configuration with disabled motion compensated temporal filtering (MCTF). All the sequences were encoded with quantization values of 22, 27, 32, 37. Then BDrate [7] is calculated for center and boundary regions using total bitrate and corresponding PSNR of those region defined in Figure 2. The overall BDrate and boundary BDrate are calculated as the average of the BDrate values among all corresponding regions. BDrate indicates the average bitrate increase for the same quality of a method with respect to the anchor method. Hence, negative value means bitrate reduction.
To gain a comprehensive understanding of the effectiveness of different methods, we design and experiment with a wide range of simulations which covers all aspects of the methods. We list all our simulations with specific naming in Table 2. In this table, the name AGCMP8+SP, for example, means aligned GCMP with pad size of eight samples and two subpictures with sharing border at discontinuity boundary.

Table 1. Test sequences

Class	Sequence name	ORG ERP Resolution@FPS	CMP Resolution	Bit depth
S1 (8K)	SkateboardInLot	8192×4096@30	6144×4096	10
	ChairLift	8192×4096@30	6144×4096	10
	KiteFlite	8192×4096@30	6144×4096	8
	Harbor	8192×4096@30	6144×4096	8
	Trolley	8192×4096@30	6144×4096	8
	GasLamp	8192×4096@30	6144×4096	8
S2 (6K)	Balboa	6144×3072@60	4608×3072	8
	Broadway	6144×3072@60	4608×3072	8
	Landing2	6144×3072@60	4608×3072	8
	BranCastle2	6144×3072@60	4608×3072	8

Table 2. List of different simulations

Simulation		Padding	Aligned	SubPic	VB
1	CMP (Anchor)	No	No	No	No
2	CMP+SP	No	No	Yes	No
3	CMP+VB	No	No	No	Yes
4	GCMP4	4	No	No	No
5	GCMP4+VB	4	No	No	No
6	AGCMP4	4	Yes	No	No
7	AGCMP4+SP	4	Yes	Yes	No
8	AGCMP4+VB	4	Yes	No	Yes
9	GCMP8	8	Yes	No	No
10	GCMP8+VB	8	No	No	Yes
11	AGCMP8	8	Yes	No	No
12	AGCMP8+SP	8	Yes	Yes	No
13	AGCMP8+VB	8	Yes	No	Yes

4.1. Objective Analysis

For the 6k (static camera) and 8k (motion camera), Figure 4 and Figure 5 shows the objective results of simulations 2, 3, 4, 6, 9, and 11 respectively. We did not include the rest of the simulations, because GCMP/AGCMP with VB or SP produces not much different results than GCMP/AGCMP. As it is readily apparent, for CMP, SP and VB give a noticeable performance gain on the discontinuity boundary, and slight gain in overall BDRate as also reported in [5]. According to Figure 4, SP works slightly better than VB for contents with static camera. The other simulations, however, show a good performance on the boundary region, with a slight increase in overall bitrate. As it can be further seen, AGCMP methods have slightly better tradeoff between boundary and overall BDRate compared to corresponding GCMP methods, especially for 6K test sequences which has static camera.

4.2. Subjective Results

To evaluate the correctness of our objective metric as well as the efficiency of our methods, in this section, we perform a subjective analysis on some of the methods. Due to page number limitation, we limit our subjective evaluation to CMP+SP and GCMP8; and compare their results against the anchor method. The decoded contents around the discontinuity border using anchor (CMP) and CMP+SP methods are illustrated in Figure 6 and Figure 7, respectively. If we slightly zoom-in on both images, we can see a seam artifact right at the discontinuity boundary of the anchor method. The artifact exists due to the information leakage of irrelevant contents between lower and upper faces during encoding. The same artifact, however, does not exist for CMP+SP output since subpicture treats discontinuity boundary as image border and does not let any information leak to the other side. This observation is completely correlated with objective result of CMP+SP, which shows a clear gain compare to the anchor. Subpicture and VB however remove the leaking artifact, they do not completely address seam artifacts related which is still visible in rendered viewport. This has been better shown in Figure 8 and Figure 9. The seam artifact is addressed in methods such as GCMP8, which produces the highest gain according to the proposed objective metric. The rendered viewport is demonstrated in Figure 10. As it can be seen, the face boundary area has become significantly smoother and less apparent compare to Figure 8 and Figure 9, which confirms the gain suggested by the proposed objective metric.

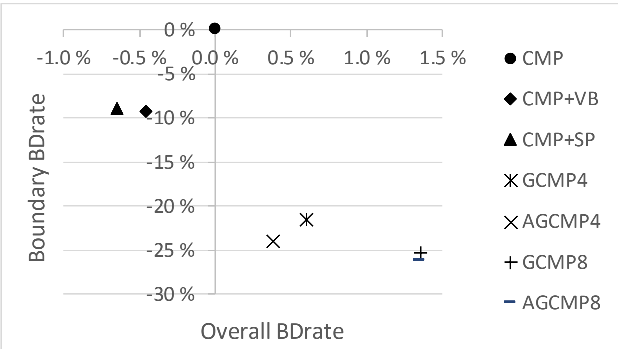

Figure 4: BDrate results for 6K test sequences (static camera).

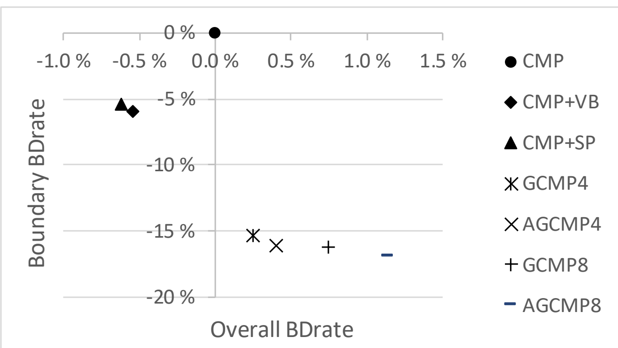

Figure 5: BDrate results for 8K test sequences (moving camera).

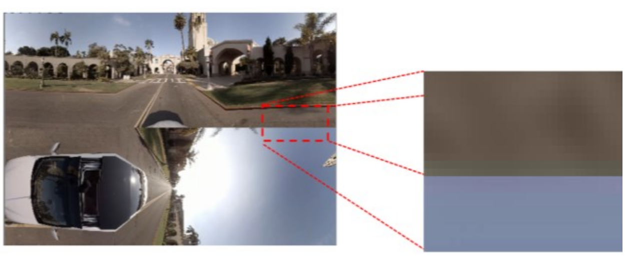

Figure 6: Anchor method output produces the greenish artifact

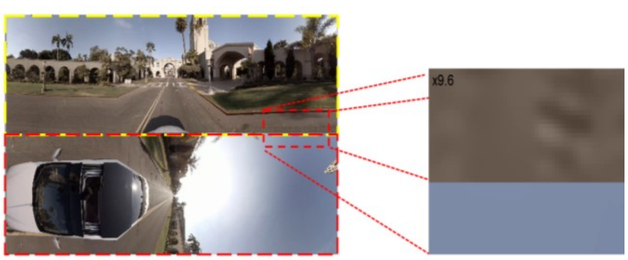

Figure 7: Leaking artifact removed under subpicture and VB

5. CONCLUSIONS

In this work, we investigated the effect of several different coding methods for CMP/GCMP format on VVC standard, among which a new technique called aligning, which reduces the amount of bitrate, while allows for combining subpicture with padding. We further introduce a new objective metric for systematic evaluation of our results in an easy and efficient way. Our metric works based on PSNR calculation of boundary regions at multiple layers. Based on this metric, so far, a simple aligning method produce the best result.

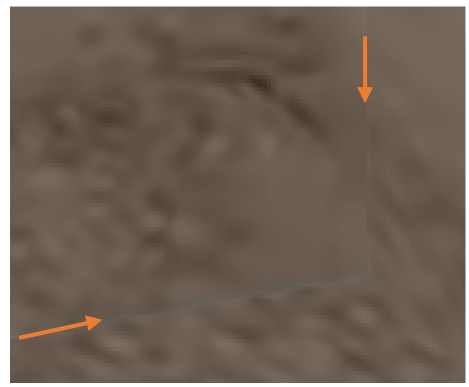

Figure 8: Artifact under anchor method

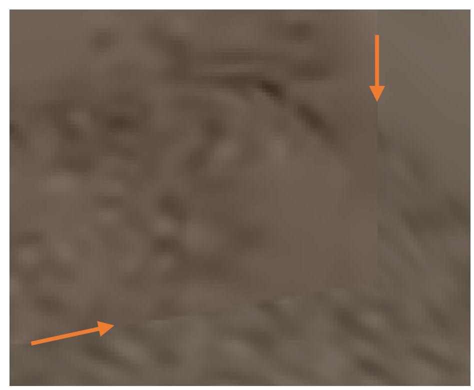

Figure 9: Subpicture (or VB) effect on seam artifact

Figure 10: AGCMP effect on seam artifact

6. REFERENCES

[1] A. Zare, A. Aminlou, M. M. Hannuksela and M. Gabbouj, "HEVC-compliant tile-based streaming of panoramic video for virtual reality applications," In Proceedings of the 2016 ACM on Multimedia Conference, pp. 601-605, Oct. 2016.

[2] Y. Sánchez, R. Skupin, and T. Schierl, "Compressed domain video processing for tile based panoramic streaming using HEVC, " Proc. of IEEE International Conference on Image Processing (ICIP), pp. 2244-2248, Sep. 2015.

[3] ITU-T Recommendation H.266, "Versatile video coding," pre-published Aug. 2020. https://www.itu.int/rec/T-REC-H.266

[4] M.Wien, V.Baroncini, "Status report on 360 video verification test preparation," JVET-T0044, Teleconference, Sep, 2020.

[5] L. Lee, J.-L. Lin, Y. Wang, Y. He, L. Zhang, "AHG6: Blending with padded samples for GCMP,", JVET-T0118, Teleconference, Oct. 2020.

[6] M. Xu, C. Li, S. Zhang, P. Le Callet, "State-of-the-art in 360 video/image processing: Perception, assessment and compression," *IEEE Journal of Selected Topics in Signal Processing*, *14*(1), pp.5-26, Jan 2020.

[7] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," VCEG-M33. 2001.

[8] Yan Ye, Jill Boyce, "Algorithm descriptions of projection format conversion and video quality metrics in 360Lib Version 9," JVET-M1004, Marrakech, MA, Jan. 2019.

[9] ITU-T Recommendation H.266, "Versatile video coding," pre-published Aug. 2020. https://www.itu.int/rec/T-REC-H.266

[10] A. Browne, J. Chen, Y. Ye, S. Kim, "Algorithm description for Versatile Video Coding and Test Model 14 (VTM 14)," JVET-W2002, July. 2021.

[11] M. Coban, R. Skupin, "AHG12: Proposed JVET common test conditions and evaluation procedures for MCTS and sub-pictures with boundary padding," JVET-M0870, Marrakech, MA, Jan. 2019.

[12] Y. He, J. Boyce, K. Choi, J.-L. Lin, "JVET common test conditions and evaluation procedures for 360° video," JVET-U2012, Jan. 2021.

[13] 360Lib-10.0. [Online]. Available: https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/tags/360Lib-10.0/

[14] VTM11.0, reference software for VVC. [Online]. Available: https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tags/VTM-11.0