INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/WG 4
MPEG VIDEO CODING

ISO/IEC JTC1/SC29/WG4 M56591
April 2021, Online

Source **Sungkyunkwan University (SKKU)**
Status **Input**
Title **[MIV] Extraction and Merging on Frame Packed Video**
Author **Jong-Beom Jeong, Soonbin Lee, Eun-Seok Ryu**

Abstract

This document provides MIV-specific tiling, merging, and frame packing methods with following experimental results. It introduces a pixel-domain and a compressed-domain frame packing to merge atlas sub-bitstreams into a single bitstream. By using the proposed packing method, one HEVC (level 6.2) decoder or multiple HEVC (level 5.2) decoders were used depending on its hardware capabilities. The experimental result showed 3.93% BD-rate gain.

1 Introduction

Proposals on frame packed video have been discussed to achieve two main goals: (1) BD-rate saving for overall quality improvement and (2) the number of decoder instances reduction for decoder synchronization [1-4]. In the proposals, the coded video sub-stream carries texture, geometry, and potentially occupancy data which are packed in frame(s) at pixel domain.

This document provides a flexible solution for packing texture and geometry, regardless of the pixel-domain frame packing. Figure 1 shows the system overview of the proposed method. Depending on the packing method, each atlas can be packed in pixel domain. Then, each video is encoded using the motion-constrained tile set (MCTS) encoder, and different tile segmentation method can be applied depending on the packing method. Each tile from the bitstream is extracted and transmitted. If the packing method requires merging, the client merges the tile bitstreams in compressed-domain to reduce the number of decoder instances. After the decoding, the frame-packed picture is unpacked.

For the proposed methods, the followings are required: tile segmentation method, tile extraction, and tile merging. This contribution introduces four MIV-specific non-uniform tile segmentation and merging methods to divide each basic views and additional views patches for quality control as described in [5], using MCTS. Tile bitstream extraction can be done using HM extractor. However, HM does not provide the tile merging software. Thus, the HM-based tile merging software has been implemented by SKKU[6] and used for this proposal.

Furthermore, because the current HM does not support individual QP allocation for each tile/slice as mentioned in [4], kvazaar, an open source HEVC encoder[7], can be used for encoding the frame-packed video. The kvazaar may provide increased bitrate compared to HM when using non-

uniform tiling because of the differences in implementation. Further, tile extraction by HM software from MCTS bitstream generated by kvazaar cannot be done. This is because HM tile extractor generates output tile bitstream by parsing extraction information set (EIS) SEI messages, and kvazaar does not generate EIS SEI messages. Therefore, HM-based tile extraction software using slice segment address has been implemented by SKKU[8] and used for this proposal.A simplified uniform frame packing and tile segmentation method which is efficient in kvazaar is explained in Section 3.

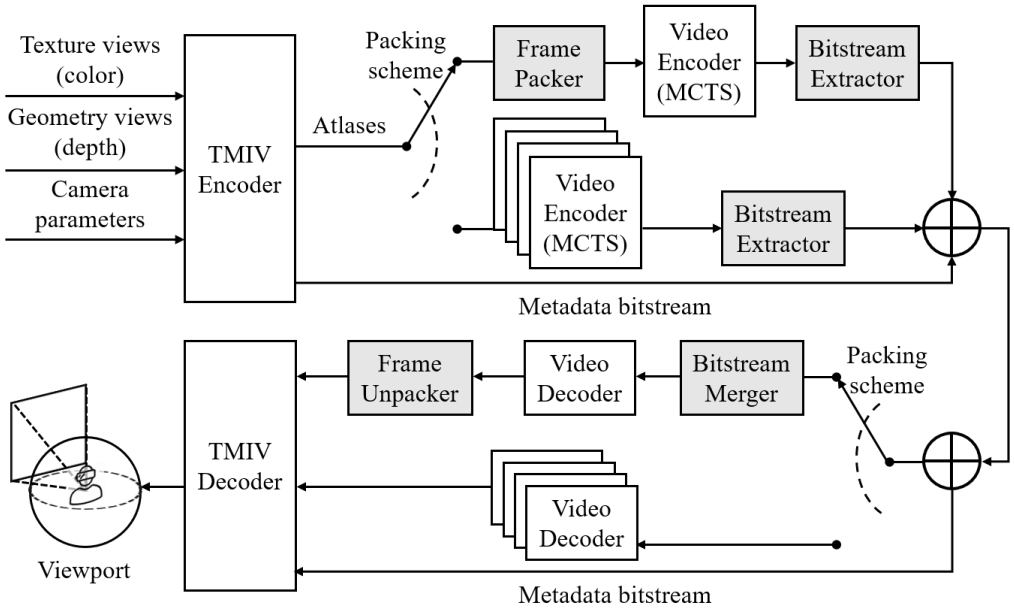


Figure 1. System overview of the proposed packing method

2 MIV-specific non-uniform tiling and packing methods

Pixel-domain frame packing has advantages on the number of decoder instances reduction. For example, four HEVC (level 5.2) bitstreams are generated in the MIV and MIV view anchors, while one HEVC (level 6.2) bitstream is generated in the frame packing. However, frame packing based MIV implementation on the legacy systems which does not has the capability to decode HEVC (level 6.2) bitstream is challenging. Therefore, a fine-grained packing method which can also cover the legacy system is needed. One of the approaches to meet the requirement is to encode the atlases using tiles, extract individual tiles, and merge them into one bitstream in compressed domain. Meanwhile, tile segmentation method was designed considering the following two requirements.

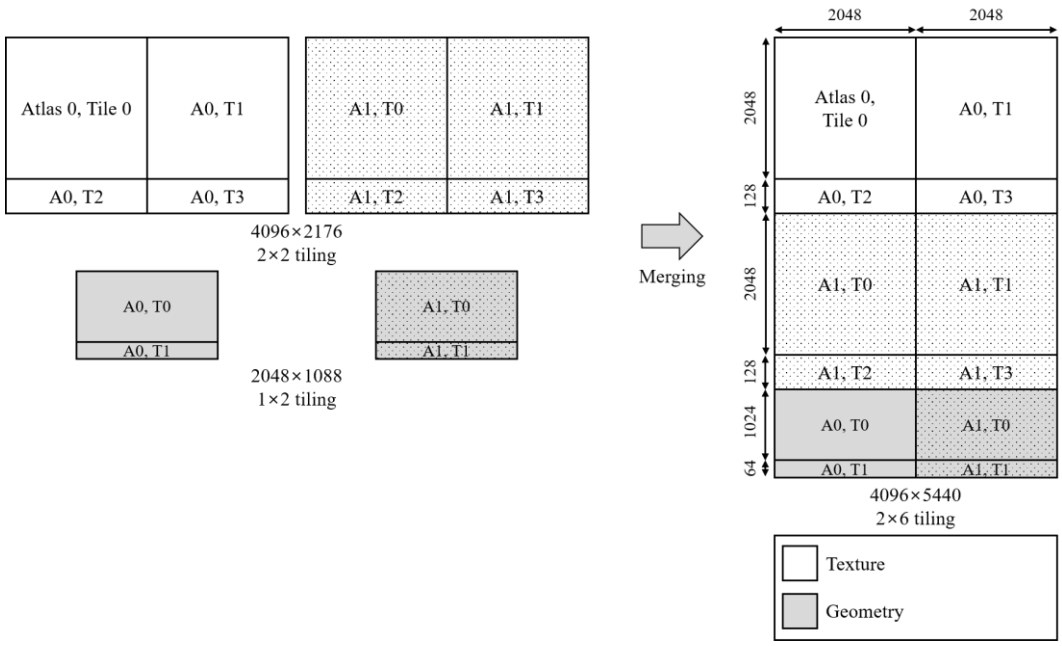
- Each basic view and additional view is distinguished by the tile boundaries, thereby an implementation of selective streaming can be made as introduced in [5].
- Tiles should be divided not to generate any empty spaces in the merged bitstream, thereby the waste of the pixel rate is not made.

Therefore, the proposed tiling and merging method are specific to the MIV test sequences. Figure 2 shows four MIV-specific non-uniform tile segmentation and merging methods on eleven MIV test sequences. Note that the resultant bitstream generated by the proposed method is compatible with HEVC/VVC decoder. Not to make empty spaces in the merged bitstream, two input tiles/slices can be located vertically in one output tile, which meets the constraints of tiles in

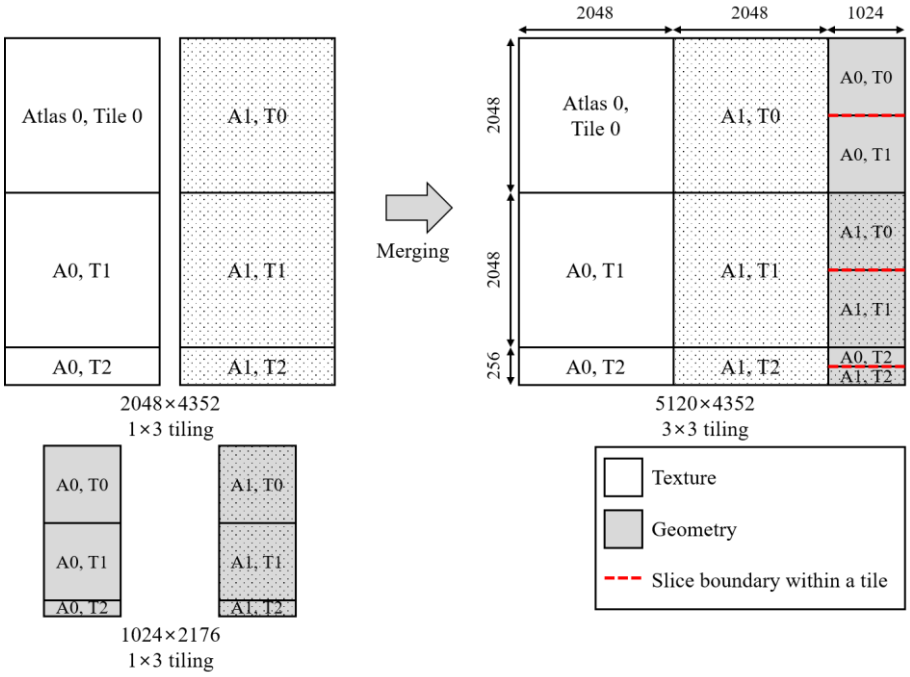
HEVC/VVC.

Experiments for this proposal were conducted following the MIV CTC[9], yet HM version 16.20 was used to encode and decode because the current VVenC does not fully support the tile and sub-picture encoding.

Mode 0, class CG-A, CG-C



Mode 1, class CG-B, CG-N, CG-Q



Mode 2, class NC-E, CG-J, CG-O, NC-P, CG-R

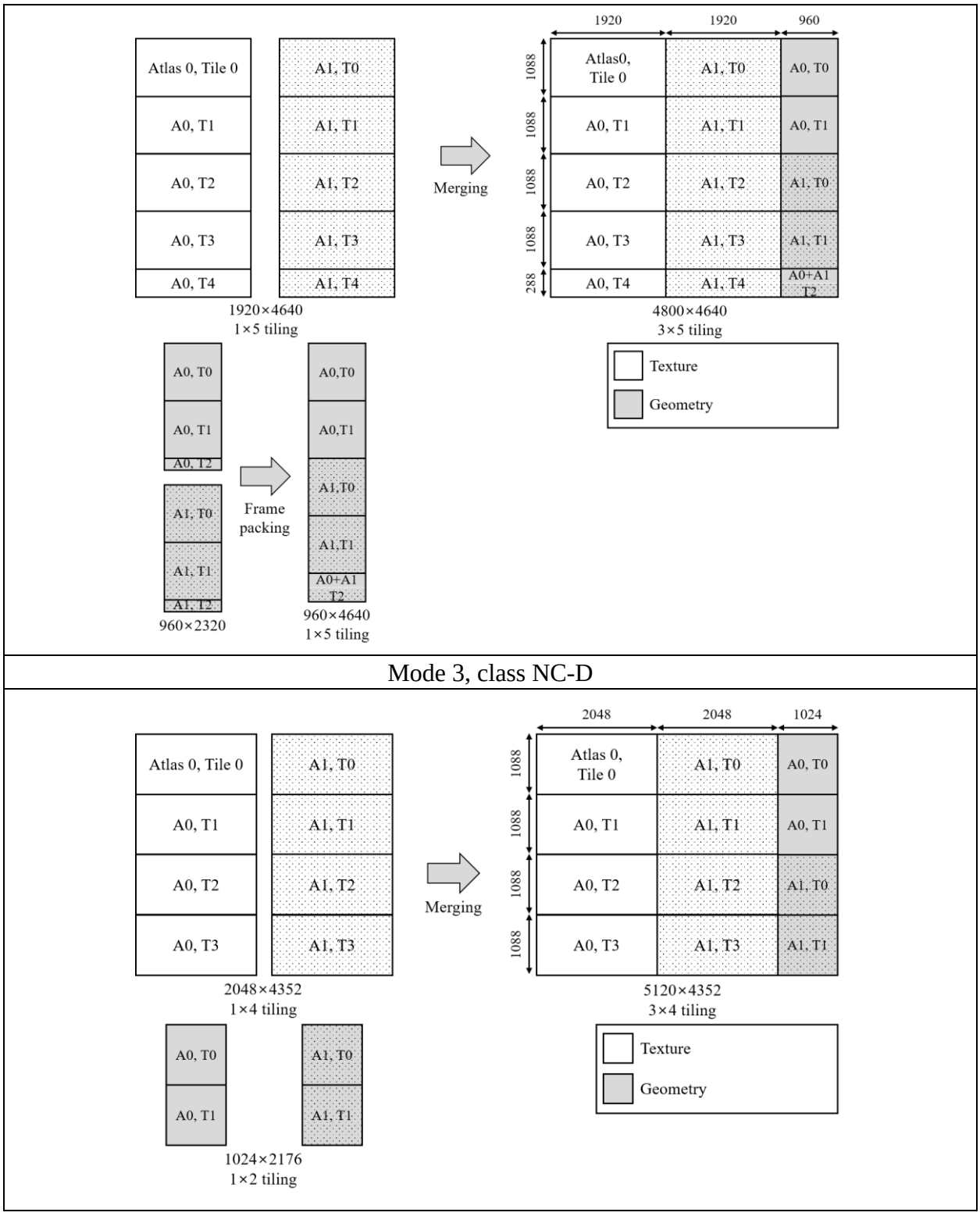


Figure 2. MIV-specific non-uniform tile segmentation and merging methods; white represents texture, gray represents geometry, dotted area distinguishes the atlas index, red-dotted lines represent the slice boundary within a tile

Table 1 below shows the experimental results of the MIV-specific non-uniform tiling and packing methods compared to the MIV anchors. Overall, the proposed method showed slight BD-rate savings. One of the reasons of this gain is deactivation of deblocking filter owing to MCTS[10].

Note that the proposed method required single decoder while the anchor needed four decoders. Otherwise, the proposed method can merge tiles and generate two, three, or four bitstreams depending on the client's capability.

Table 1. MIV-specific non-uniform tiling and packing results compared to the A17 anchors

Sequence		High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	Max delta Y-PSNR	High-BR BD rate VMAF	Low-BR BD rate VMAF	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR
ClassroomVideo	SA	0.08%	-0.70%	1.37	-1.80%	-2.23%	1.70%	0.21%
Museum	SB	-0.19%	-0.25%	16.14	0.16%	-0.25%	-0.29%	-0.38%
Fan	SO	0.06%	-0.01%	8.23	0.20%	0.16%	0.45%	0.52%
Kitchen	SJ	-0.36%	-0.05%	16.23	0.84%	-0.57%	0.09%	0.03%
Painter	SD	-0.84%	-1.30%	7.75	-0.47%	-0.81%	0.05%	-0.41%
Frog	SE	-0.46%	-0.65%	6.56	-0.42%	-0.63%	0.44%	0.25%
Carpark	SP	-1.22%	-1.55%	7.05	-0.31%	-0.71%	0.45%	0.03%
Chess	SN	-1.84%	-0.73%	14.36	-0.23%	-0.29%	-0.34%	-0.23%
Group	SR	-0.09%	-0.49%	11.71	0.19%	-0.26%	-0.19%	-0.32%
ChessPieces	SQ	-0.28%	-1.20%	14.64	-1.04%	-0.89%	1.00%	-0.36%
Hijack	SC	-0.27%	-0.77%	9.33	0.16%	-0.67%	0.12%	-0.10%
MIV		**-0.49%**	**-0.70%**	**10.31**	**-0.25%**	**-0.65%**	**0.32%**	**-0.07%**

3 Simplified uniform frame packing and tiling methods

Because the proposed method described in Section 2 showed gain on HM, experiments on the same method using kvazaar was conducted. However, experimental results showed BD-rate loss compared to the MIV anchor. Early termination and optimization of encoding on kvazaar caused different results with these of HM. By an investigation of tiling on kvazaar discovered that non-uniform tiling on kvazaar showed BD-rate loss, while uniform and simplified tiling does not. Therefore, the simplified and uniform frame packing and tiling methods for kvazaar were designed, while considering the following two requirements.

- Frame packing should not generate any empty spaces.
- Tiles should be divided to distinguish texture and geometry, to apply different QPs.

Figure 3 shows two simplified uniform frame packing and tiling schemes on eleven MIV test sequences. Note that the resolutions of each tile are same because they are uniformly tiled, while the merged two geometry atlases in the frame packed video is considered as one tile.

Experiments for this proposal were conducted following the MIV CTC[9], yet kvazaar was used to encode and decode because the current HM does not provide the individual QP allocation for each tile.

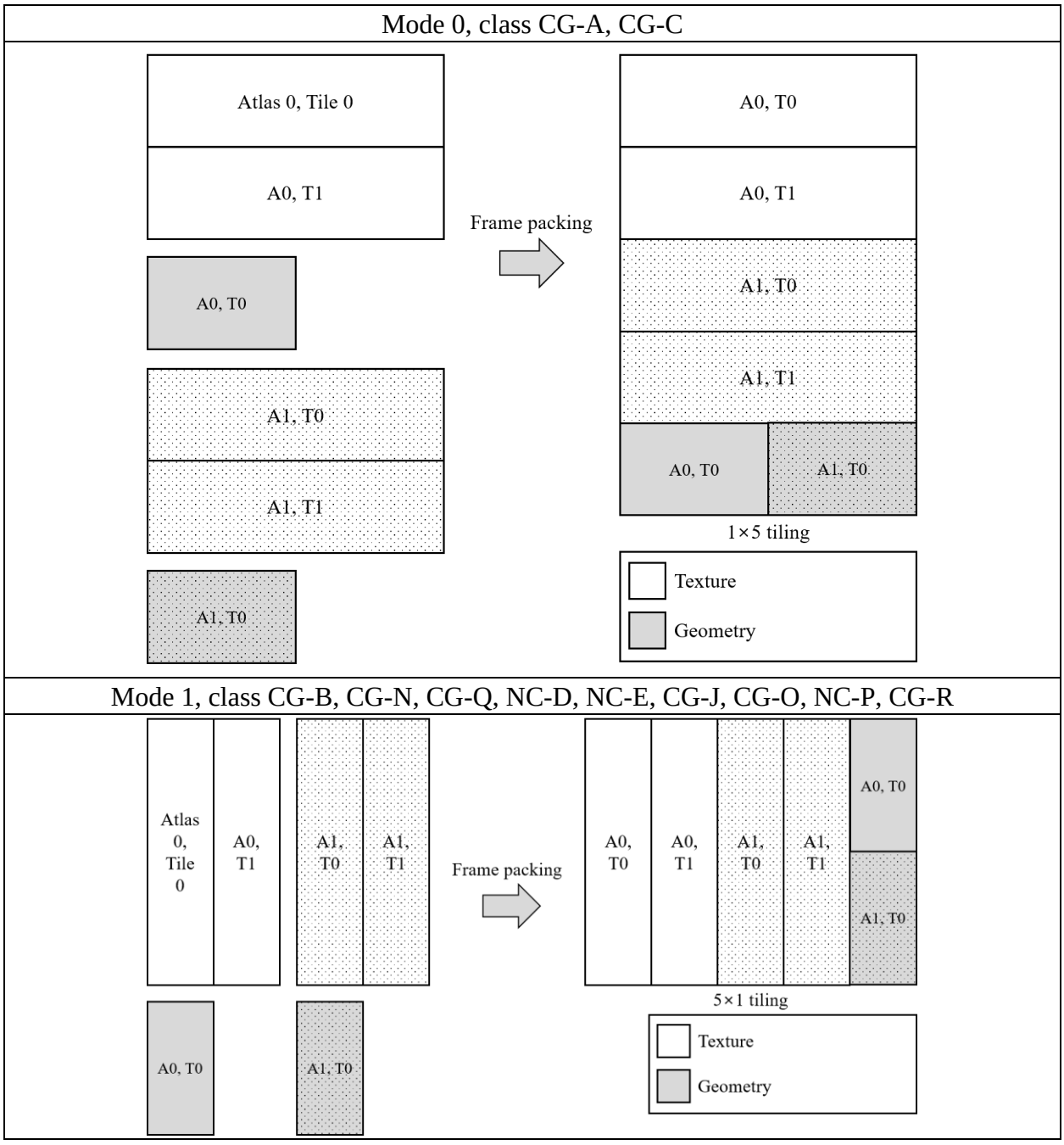


Figure 3. Simplified uniform tile segmentation and merging methods; white represents texture, gray represents geometry, dotted area distinguishes the atlas index, red-dotted lines represents the slice boundary within a tile

Table 2 below shows the experimental results of the simplified uniform frame packing and tiling methods compared to the MIV anchors generated by kvazaar. The proposed method showed 3.93% BD-rate savings on Y-PSNR. Because the proposed method used tile and they are extractable, the proposed method can generate one, two, three, four, or five bitstreams depending on the client's capability.

Table 2. Simplified uniform tiling and packing results compared to the A17 anchors

Sequence		High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	Max delta Y-PSNR	High-BR BD rate VMAF	Low-BR BD rate VMAF	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR
ClassroomVideo	SA	-1.95%	-2.39%	1.19	-7.6%	-3.7%	-3.2%	-2.9%
Museum	SB	-1.51%	-2.81%	14.94	-2.5%	-3.9%	-3.2%	-4.1%
Fan	SO	-3.76%	4.44%	7.74	-6.2%	1.8%	-6.0%	5.5%
Kitchen	SJ	-1.02%	-0.89%	15.74	-1.1%	-2.7%	-2.8%	-3.2%
Painter	SD	-12.36%	-5.81%	7.41	-13.0%	-3.2%	-12.8%	-6.4%
Frog	SE	-4.08%	-0.41%	5.93	-5.9%	-1.4%	-7.2%	-2.6%
Carpark	SP	-9.04%	-7.88%	6.81	-11.2%	-8.7%	-10.2%	-8.8%
Chess	SN	-4.78%	-4.20%	13.93	-7.1%	-6.3%	-4.2%	-5.3%
Group	SR	4.72%	1.37%	11.19	2.0%	-2.0%	-0.6%	-3.4%
ChessPieces	SQ	-8.05%	-4.24%	14.45	-8.5%	-6.7%	-7.7%	-6.4%
Hijack	SC	-1.33%	0.72%	9.04	-3.4%	0.3%	-5.1%	-1.6%
MIV		**-3.93%**	**-2.01%**	**9.85**	**-5.9%**	**-3.3%**	**-5.7%**	**-3.6%**

4 Conclusion

This document presents the MIV-specific non-uniform tiling, packing method and simplified uniform frame packing, tiling method for HM and kvazaar. Experimental results showed 0.49% and 3.93% BD-rate gains for non-uniform and uniform packing methods in high bitrate, especially efficient on NC sequences. Besides, the proposed method required single decoder instance, while having the flexibility to provide bitstreams considering the capabilities of the client.

This document recommends:

- Adding the proposed packing methods in MIV, including the tiling, extraction, merging, and decoding processes.

In the next meeting, we will provide the results of the proposed packing method on the extended profile proposed by Intel[1].

Participants who wish to obtain the tile merger[6] and the slice segment address-based tile extractor[8] may request the software coordinator listed in below.

Software coordinator:

- Jong-Beom Jeong
  uof4949@skku.edu

5 References

[1] “MIV frame packing and extended profile”, Jill Boyce, Basel Salahieh, ISO/IEC JTC1/SC29/WG11 input document m54491, July 2020, online meeting.

[2] “Improved MIV View Anchors Using Frame Packing”, Basel Salahieh, Jill Boyce, ISO/IEC JTC1/SC29/WG11 input document m55541, October 2020, online meeting.

[3] “Discussion on new packed video sub-bitstream type”, Lukasz Kondrad, Jaakko Keränen, ISO/IEC JTC1/SC29/WG11 input document m53125, April 2020, online meeting.

[4] “CE-1.3: frame packed video sub-bitstream type”, Lukasz Kondrad, Vinod Kumar Malamal Vadakital, Lauri Ilola, ISO/IEC JTC1/SC29/WG11 input document m54274, June 2020, online meeting.
[5] “[MPEG-I Visual] Report on Asymmetric Quantization on MIV”, Soonbin Lee, Jong-Beom Jeong, Eun-Seok Ryu, ISO/IEC JTC1/SC29/WG11 input document m55014, October 2020, online meeting.
[6] https://gitlab.com/MCSLJeong/hevctilemerger, accessed 04.12.2021.
[7] https://github.com/ultravideo/kvazaar, accessed 04.12.2021.
[8] https://gitlab.com/MCSLJeong/hevctileextractor, accessed 04.12.2021.
[9] “Common Test Conditions for MPEG Immersive Video”, Joel Jung, Bart Kroon, ISO/IEC JTC1/SC29/WG4 output document n00051, October 2020, online meeting.
[10] “[MPEG-I Visual] Deblocking Filter for Atlases in Basic and Additional Views”, Eun Yeo, Sang-hyo Park, Je-Won Kang, ISO/IEC JTC1/SC29/WG4 input document m55227, October 2020, online meeting.