INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11 MPEG2019/M48094
July 2019, Göteborg, Sweden

Source	**Poznań University of Technology (PUT), Poznań, Poland**
Status	**Input**
Title	**[MPEG-I Visual] High-frequency residual layer separation for immersive video coding**
Author	Olgierd Stankiewicz, Marek Domański, Dawid Mieloch, Adrian Dziembowski

1 Introduction

This document presents a technical description of high-frequency residual layer separation for immersive video coding. The method was tested with Test Model for Immersive Video using Common Test Conditions [1].

2 Overview of the proposed technique

The proposal exploits also a multi-layer approach, in which input video is splitted into layers in the spatial frequency domain. In the case of our proposal, the input video is split into two layers:

- **base layer**, which contains content that is spatially low-pass filtered, and that can be efficiently coded with classic predictive coding like HEVC,
- **residual layer**, which contains spatial high-frequency residual content.

The separation of layers occurs at the very beginning of the processing as a result of motion-compensated temporal filtering [2]. Each frame of each input view is processed independently. Block-based motion estimation is performed in order to find the motion vectors pointing to frames neighboring in time (3 previous and 3 next frames). The matched blocks are low-pass filtered. The process yields low-frequency base texture layer which is fed to the base encoder, whereas the remaining high-frequency residual part of the input video is fed to the residual layer encoder. The layer separation process is entirely automatic.

The content of the high-frequency residual layer is usually not compressed efficiently with classic predictive coding. The content of the residual layer is modeled as a non-stationary random process which can be coded jointly among the views.

The spectral envelope is estimated from the energy-normalized residual layer with the use of technique similar to LPC. The result is a set of IIR filter coefficients (in the horizontal and vertical direction) which are coded with the use of LAR (log-area-ratio) 16-bit representation. The proposed technology allows for coding of the residual layer for all views or only for one selected view. In the latter case, the residual layer in the missing views is synthesized.

After view synthesis in TMIV decoder, the high-frequency residual layer is generated synthetically (Fig. 1) basing on the spectral envelope, reconstructed basing on IIR filter

coefficients sent in LAR format. Afterwards, the high-frequency residual layer is added the content of the synthesized view.

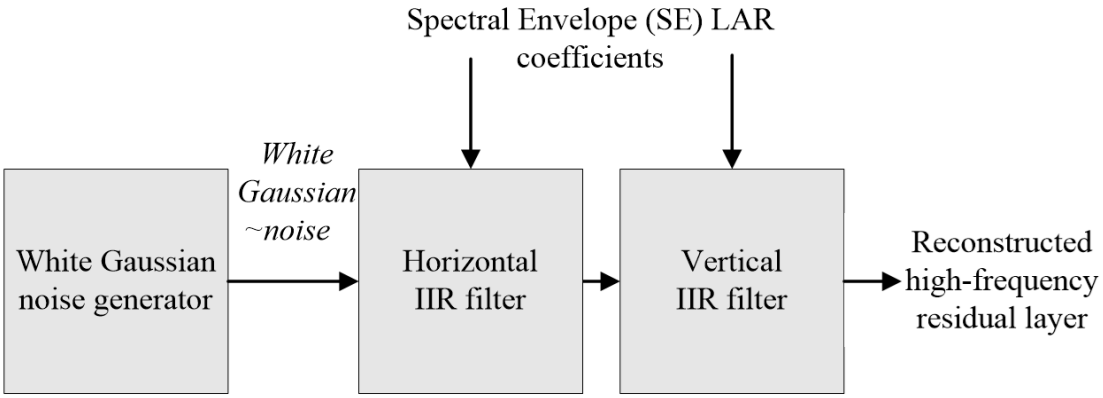

Fig. 1. Reconstruction of the high-frequency component in the decoder.

3 Experimental results

Test class	Sequence	Anchor	High-bitrate BD rate Y-WSPSNR	Low-bitrate BD rate Y-WSPSNR	High-bitrate BD rate VMAF	Low-bitrate BD rate VMAF	High-bitrate BD rate MM-SSIM	Low-bitrate BD rate MS-SSIM	Pixel rate ratio
CG1	Classroom Video	A1 (MIV anchor)	0.0%	153.6%	-24.2%	-4.0%	18.5%	17.3%	0.00%
	Technicolor Museum	B1 (MIV anchor)	0.1%	0.2%	-0.4%	-0.1%	-0.1%	0.0%	0.00%
	Technicolor Hijack	C1 (MIV anchor)	2.6%	1.1%	3.1%	1.2%	2.6%	0.9%	0.00%
		MIV anchor	0.9%	51.6%	-7.2%	-1.0%	7.0%	6.0%	0.00%
NC1	Technicolor Painter	D1 (MIV anchor)	27.1%	13.5%	0.3%	-0.8%	8.3%	12.6%	0.00%
	IntelFrog	E1 (MIV anchor)	57.1%	24.4%	-0.5%	-1.7%	53.9%	23.2%	0.00%
		MIV anchor	42.1%	18.9%	-0.1%	-1.3%	31.1%	17.9%	0.00%
All		MIV anchor	17.4%	38.5%	-4.3%	-1.1%	16.6%	10.8%	0.00%

The experimental results show that for sequences with some amount of noise (ClassroomVideo, TechnicolorPainter and IntelFrog) the proposal achieves better quality in terms of VMAF when compared with TMIV anchor. The overall bitrate of textures for abovementioned sequences is significantly smaller, e.g. in QP1 for ClassroomVideo is 5 times smaller when the proposed method is used in comparison with the anchor.

The PSNR shows a significant decrease of the quality, however, it is a result of adding the reconstructed high-frequency residual layer, based on the WG noise generator. PSNR is a point-based quality measurement, therefore, the proposed approach will always decrease the value of PSNR. In order to prove this, we calculated the PSNR for the proposed method, but without adding the reconstructed high-frequency to the synthesized view.

As it can be seen in the table below, if the synthesis uses only the base layer, then the BD-rate for PSNR is better than for the anchor. However, the subjective quality is better when the residual layer is added after the synthesis.

Test class	Sequence	Anchor	High-bitrate BD rate Y-WSPSNR	Low-bitrate BD rate Y-WSPSNR
CG1	Classroom Video	A1 (MIV anchor)	-26.5%	-9.5%
	Technicolor Museum	B1 (MIV anchor)	-1.0%	-0.5%
	Technicolor Hijack	C1 (MIV anchor)	-0.2%	-0.6%
		MIV anchor	-9.2%	-3.5%
NC1	Technicolor Painter	D1 (MIV anchor)	1.1%	-0.2%
	IntelFrog	E1 (MIV anchor)	-5.2%	-3.3%
		MIV anchor	-2.0%	-1.7%
All		MIV anchor	-6.4%	-2.8%

4 Conclusions and future work

When only the base layer is used, the PSNR BD-rate is higher, than for anchor. However, when the residual layer is added, the virtual views seem better subjectively (despite of PSNR decrease). Therefore, obtained results encourage us to perform subjective tests of the proposal before the next MPEG meeting.

5 Acknowledgement

This work was supported by the Ministry of Science and Higher Education.

6 References

[1] J. Jung, B. Kroon, J. Boyce, Common Test Conditions for Immersive Video, ISO/IEC JTC1/SC29/WG11 MPEG/N18443, Mar. 2019, Geneva, Switzerland.

[2] http://avisynth.org.ru/mvtools/mvtools.html