ADVANCED VIDEO COMPRESSION AND RENDERING FOR HIGHLY IMMERSIVE 8K+ APPLICATIONS

M. Alvarez-Mesa[1], S. Sanz-Rodríguez[1], C. C. Chi[1], M. Glowiak[2], R. Haring[3]

[1]Spin Digital Video Technologies GmbH, Germany
[2]PSNC – Poznan Supercomputing and Networking Center, Poland
[3]Ars Electronica Futurelab, Austria

ABSTRACT

In this article we present a set of tools designed to simplify the distribution, playback, and adaptation of high-quality immersive content for new immersive environments. The tools are based on a high-performance HEVC codec that allows to encode, decode, and play ultra-high-resolution video in 8K and 16K resolutions using standard computing systems. In addition, a flexible video rendering engine allows to adapt the content to different output display environments. The new tools have been validated in two advanced immersive installations, and using new immersive content created with different techniques (8K camera, 3D CGI, and point clouds) that has been produced in the context of the Immersify European Project. Initial results show that the new tools enable ultra-high-resolution immersive video playback, and first evaluations indicate that a better immersive experience for final users can be achieved, while at the same time using a simpler workflow.

INTRODUCTION

Significant advances in content production technologies, video compression, rendering, and displaying are enabling the creation of a new generation of immersive environments able to exhibit media with an unprecedented level of quality and offering a much stronger sensation of reality.

This kind of environments requires video in ultra-high resolution (including 8K and beyond), 360° and 3D video, high frame rate, and professional color formats, Sugawara and Masaoka (1). As a consequence, the uncompressed video results in huge data rates that must be distributed using advanced video codecs, Wien et al (2).

In addition, the emerging immersive environments are by design very heterogeneous, including among others: large screen visualization systems with flat and curved screens, dome projection, and cylindrical and cubic caves with large field-of-view (FoV) capabilities.

Currently, video content has to be produced for a specific target environment using distinct workflows and playback tools. This makes the production of immersive content non-reusable across different environments or increases the production costs for multi-environment productions.

In order to overcome this barrier, a set of tools has been developed where a single version of an immersive video production can be used across different environments.

All three, tools, content, and environments, have been designed and improved in an integrated approach as described in Figure 1.

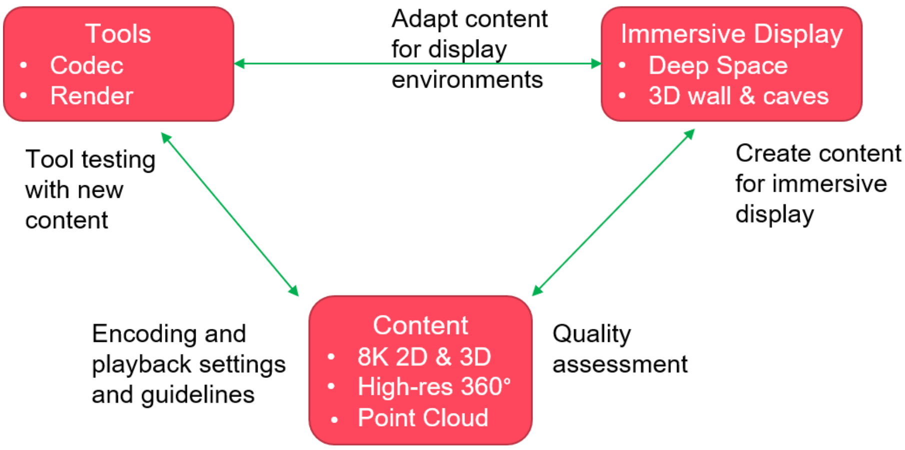

Figure 1 - Integrated approach: tools, content, and immersive display environments

The new software has been validated on two immersive environments: The Deep Space 8K at Ars Electronica Center in Linz (Austria), and the 8K 3D video wall at the Poznan Supercomputing and Networking Center in Poznan (Poland).

Moreover, in order to test the new tools in the new environments, experimental content has been produced in 8K+ resolution using very recent technologies in media acquisition such as 8K 2D and 3D camera footage, high-resolution 360° video, time-lapse photography, cinematic rendering, as well as point cloud rendering of laser scanning.

NEW IMMERSIVE ENVIRONMENTS

Immersive spaces are an alternative way to present immersive media targeting groups audiences; they include multi-display setups in museums, new immersive cinemas, and dome projection. Although these immersive display environments are highly heterogeneous, they share a common set of requirements and technical capabilities:

- **Very high-quality and high-resolution video**: As immersive environments by definition try to make the user completely involved in the content, they tend to have bigger screens with higher resolution in order to provide a higher FoV compared to TV or cinema. The requirements translate in the need of resolutions beyond current TV formats, including 8K and 16K.
- **3D (stereoscopic) video**: Most of the new immersive environments support stereoscopic content. As a result, some environments require resolutions and frame rates as high as 8Kx8K at 120 Hz (60 Hz per eye).
- **Immersive audio**: Also called 3D or spatial audio, it enhances the immersive experience by involving the audience in audio coming from all (3D) spatial directions.

Below we present a description of two specific immersive environments that have been used in the design and validation of the tools presented in this paper.

Deep Space 8K at the Ars Electronica Center

Deep Space 8K is a multifunctional presentation room located at the Ars Electronica Center in Linz. It consists of a 16 by 9 meters wall and a 16 by 9 meters floor projection (see Figure 2). It offers a wide range of artistic, scientific and pedagogical projects such as gigapixel photography, time-lapse videos, interactive experiences and many more.

Figure 2 - Program display at Deep Space 8K

Eight 4K projectors are used to generate up to two 8K stereoscopic images for the wall and the floor with a framerate of up to 120 Hz (60 Hz for each eye). Actual resolution varies depending on geometric calibration. The total resolution and frame rate of the system when taking into account both the floor and wall projection combined is up to 8Kx8K at 120 Hz.

8K 3D wall in New Media Laboratory at PSNC

The 8K 3D 60p wall installed in New Media Laboratory is composed of 12 Barco projectors in rear projection configuration and it is supported by Barco processing system based on “e2” and “MCM-50 units”. The general architecture is depicted in Figure 3. The dimensions and resolutions of the installation are 6 by 2.8 meters, and total resolution is 8192 x 4320 pixels at 120 Hz.

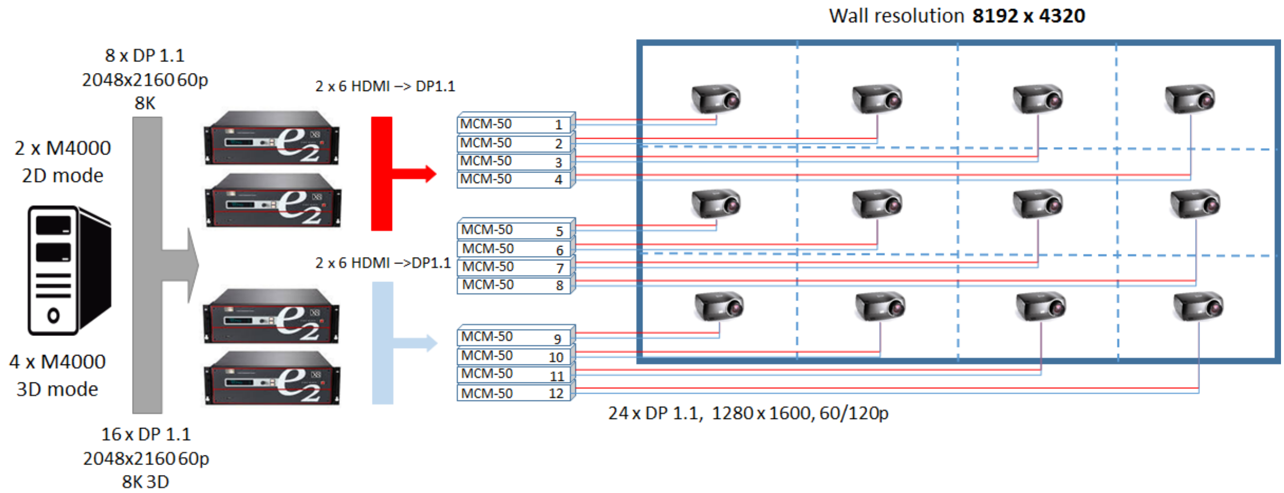

Figure 3 - General architecture of the 8K 3D rear projection wall at PSNC

The visualization system supports 3D content, with four BARCO “e2” processors grouped by two systems, each pair is stacked and works as a single virtual device serving 8K image. So, each pair of devices processes one 8K image - one for left and one for right eye, Kurowski et al (3).

IMMERSIVE CONTENT CREATION

Currently there is not a single standard way of creating immersive content, especially when targeting immersive environments. Content creators are playing with a variety of tools and techniques including:

- **Time-lapse photography**: High-resolution photo cameras are used to create time-lapse videos generating a high sensation of reality as they can go beyond the resolution of video cameras. Some examples of high-quality time-lapse video in 8K and 10K resolution include the works of Martin Heck from Timestorm Films, Joe Capra from Scientifantastic, and NHK Technologies.
- **Real-life 8K footage 2D and 3D**: By using 8K video cameras it is possible to create a more immersive content experience. 8K content has been pioneered by NHK in Japan, and new production centers are appearing including the 8K Media Laboratory of PSNC in Poland.
- **3D CGI and animation**: Computer-generated content in very high resolution allows to explore artificial worlds with a higher sense of detail and immersion. Recent examples of special use of CGI in very high resolution and immersive formats include: *Prima Materia 8K* by Nohlab, in 8Kx8K 3D; and *Singing Sand* by Tadej Droljc, in 8Kx4K 3D.
- **Panoramic video**: 360° video is being produced with high-end cameras that allow for a more immersive experience than is currently possible with head-mounted displays (HMDs). For example, the HHI Omnicam when combined with ARRI Alexa cinema cameras has been used to produce content with a FoV of 360°x120° and a resolution of 14Kx2K.
- **3D laser scanning and point cloud rendering**: In this method a scanner is used for shooting a laser beam and collecting measured distance in every direction. The collected data can be represented as points in three-dimensional space also called point clouds, representing a 3D model of complete objects. Point clouds can be rendered to 2D and 3D video using dedicated software such as CloudCompare. We extended this tool to support high-resolution 360° video rendering. Recent examples of point cloud rendering used for immersive video production include: *The Great Pyramid in 3D* by BBC Studios and Scanlab in 12Kp60 3D 360° format; and *From the Inside – A different view of the Cathedral in Poznan* by PSNC and rendered in 16Kp60 360° format.

Several content pieces used the techniques mentioned above has been produced within the context of the Immersify European Project, and are is being used for testing the tools and evaluating the immersive environments, Glowiak (4).

WORKFLOW AND TOOLS FOR CODING, AND PLAYBACK OF IMMERSIVE MEDIA

As immersive content is being produced using a wide range of techniques and it targets a wide range of heterogeneous immersive environments, it is desirable to define a workflow with a set of defined formats and tools for playback that adapt the content to the display environment. This proposed workflow has four main components: video coding, video playback, immersive audio, and media adaptation.

Video Coding

As immersive media is produced in very high resolution and quality the resulting uncompressed files are huge, as an example: 8Kx8K at 120 Hz results in a data rate of 212 Gbit/s, equivalent to 1.6 TB/min. Video compression is needed in order to make distribution and playback practical and cost effective. We selected a software implementation of HEVC as the main codec for immersive media as it offers high compression efficiency with high quality, and the flexibility to support various chroma formats and resolutions.

As a baseline we selected an existing HEVC encoder developed by Spin Digital that already had support for HEVC Range Extensions and exhibits good compression and quality performance compared to existing codecs. The codec has been enhanced with optimized pre-processing filters for geometry conversions: Equirectangular projection (ERP) to Cubemap (CMP) projection, viewport generation, and zoom and rotation functionalities. Video color adaptations including color gamut mapping, and HDR tone mapping, are supported as well. As a result, it is possible to process files coming from high-end 360° workflows, such as the Omnicam, perform color and geometry conversions, and encode in high-quality HEVC format. The encoder was also enhanced with support for parallel tile encoding and motion-constrained tiles, which are needed for applications requiring partial decoding. Table 1 summarizes the formats supported by the enhanced HEVC encoder.

Feature	Input	Output
Max resolution: 16Kx8K	RGB, 10-/12-bit	4:2:0 10-bit
360° video	ERP	CMP
Codec/format	DPX, ProRes, DNxHD	HEVC
File format	DPX, MOV	MP4

Table 1. Formats supported by the Spin Digital HEVC encoder

Experimental results show that the enhanced encoder is able to achieve a bitrate reduction of 27% for 8K video at the same objective quality and encoding speed compared to the open-source encoder x265.

Video Decoding and Rendering

On the playback side, several optimizations have been applied to a software HEVC decoder and a video render engine in order to guarantee real-time playback for very high-resolution video. These optimizations include:

- **Advanced multi-threading**: The HEVC video software decoder has been extensively optimized for modern CPU architectures that have tens to hundreds of processors cores. The HEVC decoder parallel processing works best with the Wavefront Parallel Processing (WPP) approach, as it exhibits high parallel scalability, and allows to reduce memory usage and decoding latency.

- **SIMD optimization**: The decoder has been optimized with the latest Single Instruction Multiple Data (SIMD) instructions such as Intel AVX-512. Experimental results show a performance gain of up to 20% compared to the optimized baseline with AVX2 instructions.
- **Efficient pixel formats**: Conventionally pixels or video samples are stored in computer memory in multiples of 8-bits. For the 10- and 12-bit video, the samples are normally stored in 16-bits. Compared to 8-bit, in practice 10- and 12-bit doubles the required memory and PCIe bandwidth. To alleviate this, we introduced bit aligned formats where the samples are stored directly in memory one after another without byte alignment. These formats can save up to 37.5% on required bandwidth and memory for 10- and 12-bit respectively, Alvarez-Mesa and Chi (5).
- **BC4 compression**: For some applications (e.g. 16K) even with the bit-aligned formats the main bottleneck is still the CPU to GPU data transfer. To address this issue, we added on-the-fly compression to a GPU texture format that offers up to 4x times compression over the planar format. The video renderer was able to speed up significantly, with performance gains inversely proportional to the transfer reduction.

Overall, the decoding and rendering optimizations results show that it is possible to decode and render up to 16Kx8K 60 Hz video on a latest-generation dual-socket workstation. This enhanced performance enables new applications for immersive media playback such as 8Kx8K 120 Hz stereoscopic content, very high-resolution 360° video, and very high-bitrate content (e.g. point cloud).

Table 2 shows the different immersive formats under consideration and the corresponding uncompressed and HEVC compressed bitrates for high-quality playback. The content can be played using the optimized software on a PC system with a dual socket Intel Xeon Platinum 8168 CPU (2x 24 cores) and an AMD Radeon Pro WX7100 GPU.

Format and application	Uncompressed bitrate (RGB – 10-bit)	HEVC bitrate (4:2:0 – 10-bit)
Omnicam 360°: 12Kx8K 2D 24 Hz	68 Gbit/s	80 Mbit/s
Deep space 8K: 8Kx8K 3D 60Hz	212 Gbit/s	400 Mbit/s
Point cloud 8K: 8Kx4K 2D 60 Hz	60 Gbit/s	1600 Mbit/s

Table 2 - Immersive formats and their uncompressed and HEVC compressed bitrates

It should be noted that HEVC is not optimized for point-cloud content, and as a result very high bitrates are needed for high quality playback. There are on-going efforts under MPEG for the development of a specific video codec for point clouds that will allow to compress directly the point cloud representation, Schwarz et al (6).

A different solution for reducing the bitrate without using point cloud representations includes the use of specific tools for screen content coding (SCC), as defined in the HEVC SCC extension, Xu et al (7).

Immersive Audio

Spatial sound is a fundamental component to make the final user really involved in the immersive media content. The most popular spatial audio formats include: multi-channel, objects, and ambisonics, Nachbar et al (8).

Unlike multi-channel or objects, the ambisonics format does not carry loudspeaker or object signals, but channels representing the sound field in the space. This property makes ambisonics a flexible audio format for heterogeneous immersive spaces, since one single file can be decoded to different loudspeaker layouts as well as binaural sound for VR HMDs.

The proposed workflow for immersive audio coding and playback is summarized next. On the encoder side, a high-order ambisonics (HOA) file with up to 16 channels in ACN/SN3D format, Nachbar et al (7), is encoded in Advanced Audio Coding (AAC), and then encapsulated together with the HEVC video in MP4. On the playback side, the HOA AAC track is decoded to a particular loudspeaker layout or headphones. This workflow requires the implementation of ambisonics decoders for each target loudspeaker layout.

Adapting Content to Display

Immersive content needs to be adapted to the specifics of each immersive environment. Adaptations can include: color conversions and geometry adaptations.

Color video adaptations

With the emergence of wide color gamut and High Dynamic Range (HDR) new content is currently being produced in these formats including the use of BT.2020 color gamut and HDR transfer functions (PQ and HLG). Few installed immersive display systems, however, support the new formats. In order to have a high-quality content-to-display adaptation, we implemented tone mapping (HDR to SDR) and color gamut mapping (BT.2020 to BT.709) algorithms. They are available as pre-processing filters before encoding for off-line conversion, or as post-processing filters in the video rendering module of the media player for real-time conversion.

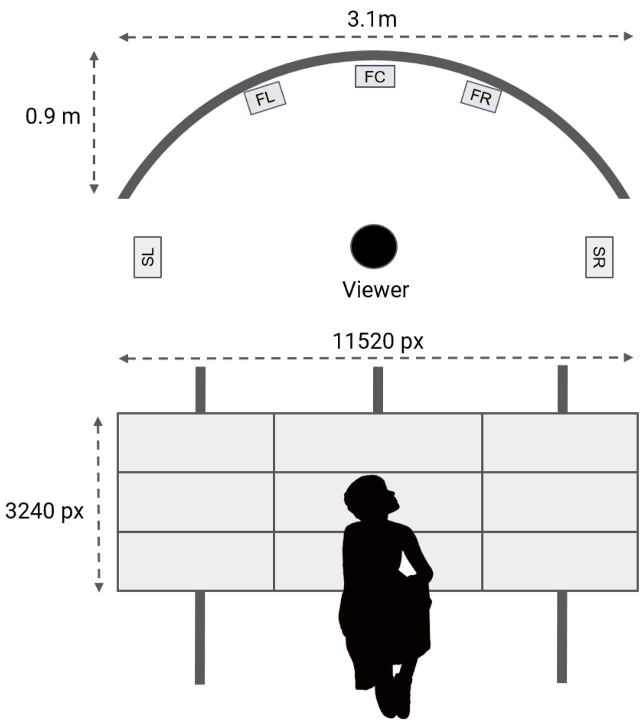

Figure 4 – Multi-screen curved display in 11520x3240 resolution

Geometry and display adaptations

The objective of these adaptations is to allow the playback of a single input file on different display environments such as high-resolution flat and curved screens. A workflow for 360° video is as follows. The starting point is an uncompressed ERP high-resolution 360° video. This file is converted, before encoding, to CMP format for better compression performance and easy of playback. Depending on the target viewing environment, adapted viewports with different viewpoints, FoVs, and resolutions can be extracted and then projected onto either regular flat or curved displays. Geometry conversions can also be applied either before encoding or during playback depending on the requirements of the target system.

Figure 4 shows an example of multi-display curved environment. It consists of 3x3 ultra-wide curved monitors (3840x1080 pixels at 144 Hz) for a total resolution of 11520x3240 pixels and a horizontal FoV of 180°, and a 5.1 surround sound system. This configuration has been used for a demonstration of immersive media at the Cannes XR section of the Marché du Film - Festival de Cannes 2019, Kuthan (9).

CONCLUSIONS

A set of tools designed for the playback and adaptation of high-quality immersive content for new immersive environments have been presented. They are based on an optimized HEVC codec implementation that allows to encode and play very high-resolution video (8K and 16K) using standard computing platforms. A set of video processing filters and a flexible video rendering engine has been designed for simplifying the adaptation of immersive content to different types of immersive displays. They include conversions such as tone and gamut mapping, 360° projection conversions, and geometry transformations. Initial results show that is possible to encode and play ultra high-quality video including 8Kx8K 3D (120 Hz), 12Kx8K 360°, and 8Kx4K point clouds. Next steps in this research include the addition of video and audio interaction capabilities for group environments, and the support of real-time geometry conversion for multi-display environments.

ACKNOWLEDGMENTS

This activity has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 762079 (www.immersify.eu).

REFERENCES

1. Sugawara, M. and Masaoka, K., 2013. UHDTV Image Format for Better Visual Experience. Proceedings of the IEEE. January 2013, vol. 101, issue. 1, pp. 8-17.

2. Wien, M., Boyce, J. M., Stockhammer, T. and Peng W., 2019. Standardization Status of Immersive Video Coding. IEEE Journal on Emerging and Selected Topics in Circuits and Systems. March 2019, vol. 9, no. 1, pp. 5-17.

3. Kurowski, K., Glowiak, M., Ludwiczak, B., Strozyk, M., Ciżnicki, M., Binczewski, A. and Alvarez-Mesa, M., 2018. PSNC advanced multimedia and visualization infrastructures, services and applications. Proceedings of the 23rd International ACM Conference on 3D Web Technology. June 20 - 22, 2018, Poznan - Poland, Article No. 34, pp 34:1-34:2.

4. Glowiak, M., 2019. Report on Immersive Content Production. Immersify project website. https://immersify.eu/wp-content/uploads/2019/03/Deliverable_D5.1-Report_on_Immersive_Content_Production_Part_1.pdf

5. Alvarez-Mesa, M. and Chi, C. C., 2018. HEVC Software Media Player for Ultra-High-Quality Video: 8K and Beyond. Proceedings of the 25th International Display Workshops. Nagoya - Japan, 2018. pp. 1437-1440.

6. Schwarz, S. et al., 2019. Emerging MPEG Standards for Point Cloud Compression. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, March 2019. vol. 9, no. 1, pp. 133-148.

7. Xu J., Joshi R., and Cohen R. A. Overview of the Emerging HEVC Screen Content Coding Extension, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 26, Issue 1, pp. 50-62, Sept. 2015.

8. Nachbar, C., Zotter, F., Deleflie, E. and Sontacchi, A., 2011. AmbiX - A suggested Ambisonics Format. Ambisonics Symposium 2011. June 2-3, 2011, Lexington (KY) - USA.

9. Kuthan, A., 2019. Immersify at Marché du Film – Festival de Cannes 2019. Immersify project website. https://immersify.eu/news/immersify-at-marche-du-film-festival-de-cannes-2019/