Fraunhofer HHI

August 2018

Live Streaming of Ultra-High Resolution 360-Degree Video with MPEG-OMAF and HEVC Tiles

This article shows how Ultra-High resolution 360-degree video captured with the OmniCam-360 can be streamed live to mobile devices using the MPEG-OMAF standard in combination with HEVC compression.

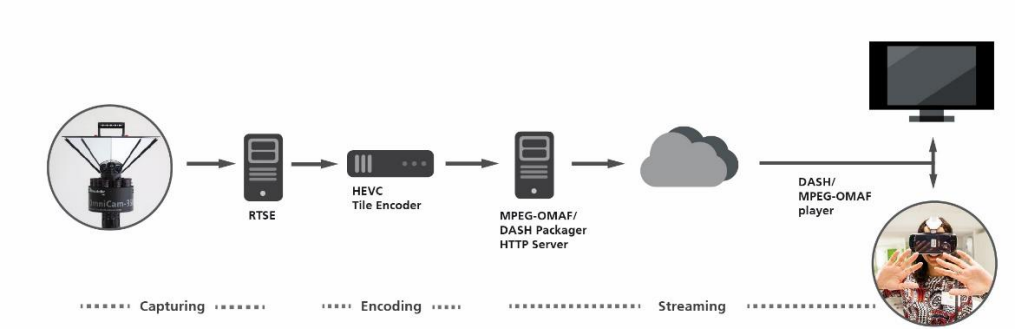

Figure 1: Continuous Live Streaming of VR360-Degrees 10K-Video with MPEG-OMAF and HEVC Tiles

The recently adopted MPEG-OMAF standard allows a significant increase in the quality of 360-degree videos for Virtual Reality (VR). By combining with HEVC, 360-degree videos with resolutions of 6K x 4K can be transmitted to the latest mobile devices, thereby taking full advantage of the native resolution of the mobile display. The video is segmented into individual tiles and each tile is encoded in its original resolution as well as in a reduced resolution. The receiver selects the tiles in accordance to the visual field in such a way, as that it ultimately transmits the high-resolution tiles to the user's viewport and the low-resolution tiles for the areas out of sight. At the end device, e.g. VR-glasses or TV set, the received tiles are combined to a HEVC-compatible video stream and can be decoded with any 4K-compatible HEVC decoder. At IBC 2018, Fraunhofer HHI will present the implementation of the new standard in a continuous real-time transmission chain. Figure 1 illustrates the complete live-chain with the following key components:

Figure 2: OmniCam-360

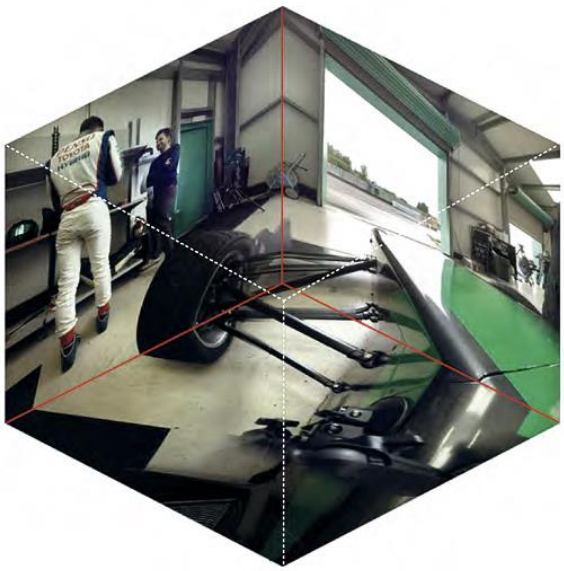

Figure 3: Example of a Cubemap projection

Capturing:

The OmniCam-360 from Fraunhofer HHI is a system for real-time recording of high-resolution 360-degree videos with a significantly higher resolution than 4K. Using the Real-Time Stitching Engine (RTSE), the eleven individual camera segments are seamlessly combined without artifacts to form a high-resolution VR360 video.

Figure 2 shows a photo of the current Omnicam-360. Ten HD cameras and one UHD camera are used to take pictures with a panorama resolution of approx. 10.000 x 3.300 pixels at an opening angle of 360 x 120 degrees. The cameras used here support 25, 30, 50 and 60 frames per second and have a global shutter and gunlock to ensure perfect stitching of the individual camera segments even under critical conditions. The fourteen HD signals are processed into a seamless panorama in real time with a latency of five frames.

For display or further processing, the resulting panorama can be transferred both in Equirectangular format and in Cubemap format, though the Cubemap format is more favorable for HEVC Tile Based Streaming with MPEG-OMAF. In Cubemap projection, as shown in Figure 3, the 360-degree view is projected onto the six inner sides of a cube, each with a resolution of 2.048 x 2.048 pixels.

A plug-in for the post-production environment Nuke/CARA, which was specially developed for the OmniCam family, is available for calibrating the camera system. It automatically generates the correction parameters for live stitching of the individual cameras and also allows the signals to be processed later in post-production.

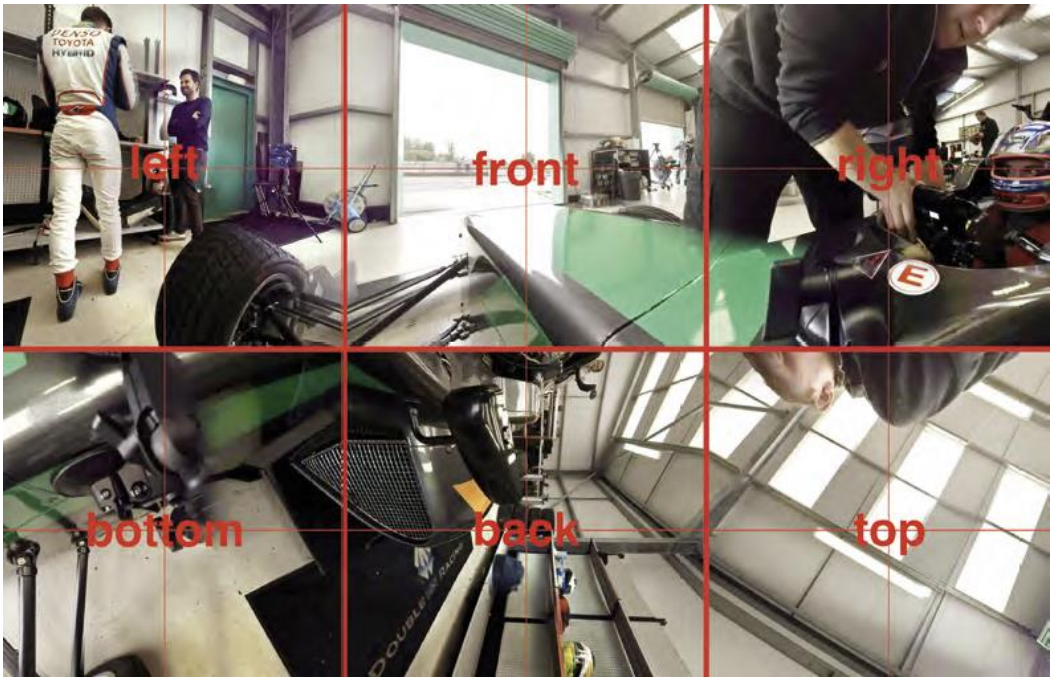

Figure 4: Segmentation of the Cubemap projection tiles

Encoding:

The latest version of the software-based live encoder from Fraunhofer HHI integrates HEVC tiles for efficient encoding and transmission of 360-degree video. The video of the OmniCam-360 is segmented into tiles and each tile is encoded separately with HEVC. In order for the tiles to be combined into a common bit stream of different qualities on the end device, the individual tiles must be able to be decoded independently of one another. This is achieved by limiting the motion vectors for inter-picture prediction at the edges of the tiles as supported by the HEVC standard.

The RTSE of the OmniCam-360 delivers the 360-degree video as a Cubemap projection in YUV 4:2:0 planar format over 10 Gbit/s Ethernet. Each of the six sides of the cube corresponds to one of the fields of view: Front, Left, Right, Back, Top and Bottom. A three times two arrangement of the six sides (2048 x 2048) results in a resolution of 6.144 x 4.096 pixels. Before encoding, these cube sides are split into four 1.024 x 1.024 tiles, resulting in a total of 24 tiles as shown in Figure 4. Each of the 24 tiles is encoded live in two different resolution levels, once in the original and once in reduced resolution at 512 x 512.

Two servers, each with two Xeon E5-2697v3 14-core CPUs, are used for HEVC encoding. Figure 5 shows the block diagram of the HEVC tile encoder. Server 1 segments the Cubemap images of the OmniCam-360 into 24 tiles, scales down the resolution to 512 x 512 pixels and then encodes these low-resolution tiles. Server 2 encodes the 24 tiles of the highest quality with the original resolution of 1.024 x 1.024 pixels per tile. The delay of the encoder in the NAB demonstrator is 2 seconds.

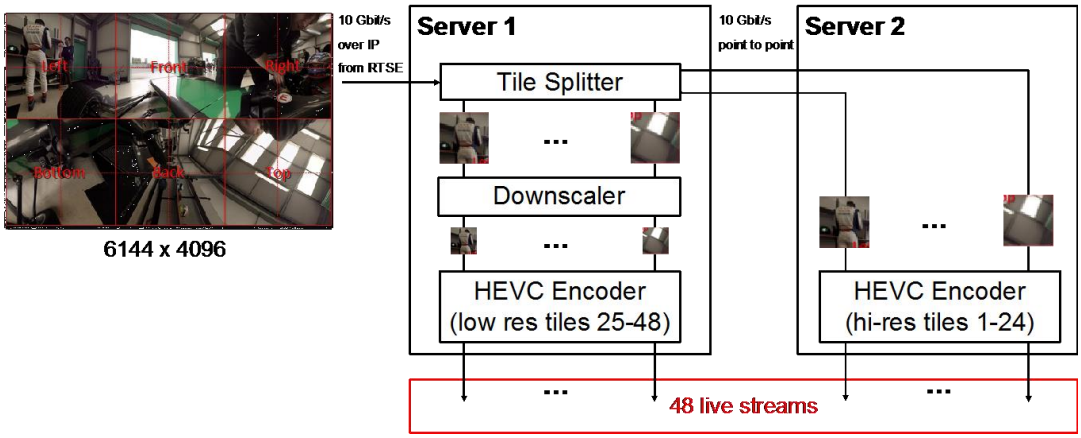

Figure 5: Block diagram of the HEVC Tile-Encoder

Streaming

The Viewport-Dependent Media Profile of the MPEG-OMAF standard is used for streaming. The DASH Packager receives the 48 live encoded HEVC streams, then synchronizes and cuts them to the key frames. Then each segment is packaged individually and made available as a DASH segment on a simple HTTP server, creating 24 "Extractor-Track" segments (one for each view port) in addition to the 48 tile stream segments. The Extractors describe for each viewport which tile stream segments belong to which specific viewport. In addition, the initialization segments and a DASH manifest with all necessary OMAF metadata are generated once in the beginning. The upper part of figure 6 illustrates how the OMAF/DASH packager works.

Since all separately encoded HEVC tiles are now made available on the server in two resolutions in the form of DASH segments, each receiver, such as VR glasses or television sets, can download the suitable HEVC tiles with the high resolution for the current viewport position. Outside the viewport, the HEVC tiles are consequently downloaded in a low resolution. OMAF metadata are analyzed so that each receiver can independently determine which tiles exactly are to be downloaded in high resolution. "Sphere region quality ranking" (SRQR for short) signals the region-dependent quality in 3D space and its assignment to respective DASH segments (tiles). The exact function of an OMAF player is illustrated in the lower part of Figure 6 and is briefly explained below.

After the respective 24 DASH segments have been downloaded for the entire 360-degree image, these are aggregated together with an Extractor Track segment and the corresponding Initialization segment to a file and passed to the mp4 parser. This parser extracts a single HEVC bit stream from the file, as specified in the extractor track, which is passed to the decoder. The decoded image is then projected onto the individual regions of the cube using OMAF region-wise packing and projection format metadata. The high-resolution tiles land on the cube surfaces in the viewport and the low-resolution tiles are scaled up and distributed on the remaining surfaces. This achieves a

higher resolution in the viewport, which would not be possible with a traditional method.

To ensure that users can always see the high-resolution video even when the viewport moves quickly, the segments are prepared in such a way that a switching process can be carried out with every new frame (0.36 seconds). Depending on the current viewport, the bit rate of the transmitted content is 5 to 10 Mbps.

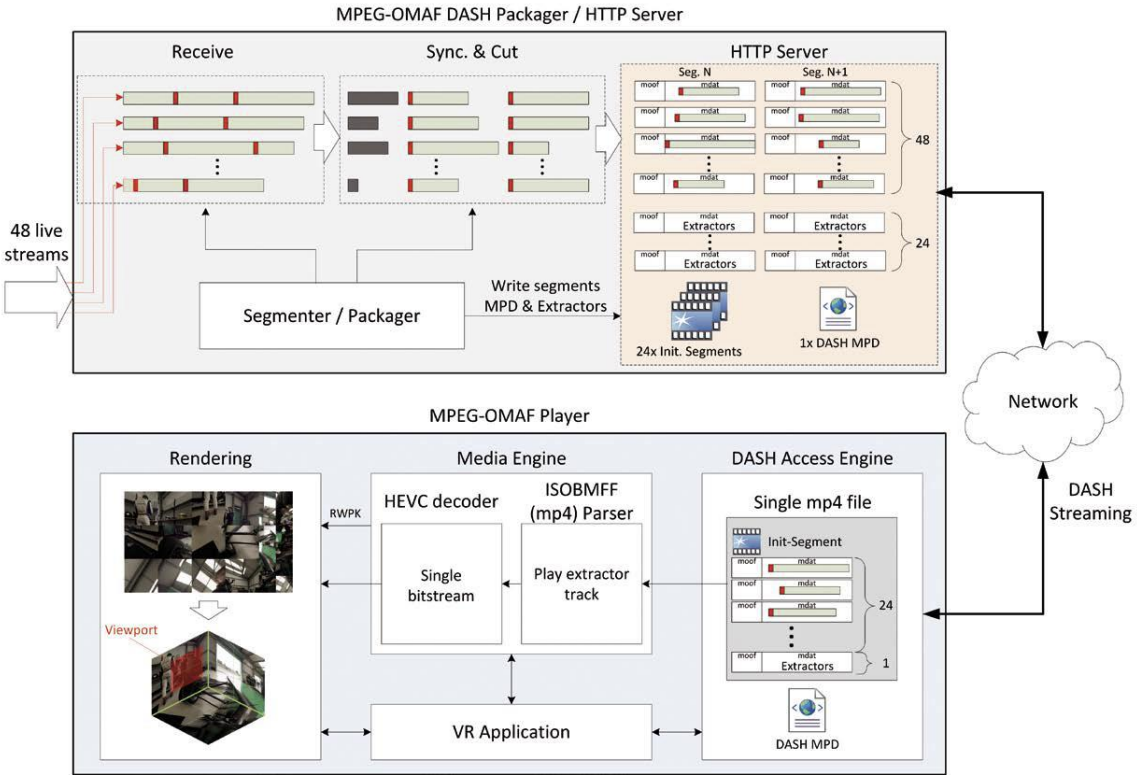

Figure 6: Block diagram of the OMAF packager and player

Summary

With the combination of MPEG-OMAF and HEVC tiles, there is now an international standard for the distribution of high-resolution 360-degree videos on existing mobile devices using existing video decoders. The demonstrator presented at NAB shows the potential of the new MPEG-OMAF standards for 360-degree videos. A complete and standards-compliant implementation of the entire transmission chain was demonstrated. The increased resolution in the mobile headset has also significantly improved the visual quality. Impressing is also the total latency of the live system from the OmniCam to the VR glasses of a total of six seconds.

Contact: cornelius.hellge@hhi.fraunhofer.de