LESSONS LEARNT DURING ONE YEAR OF COMMERCIAL VOLUMETRIC VIDEO PRODUCTION

O. Schreer[1], I. Feldmann[1], P. Kauff[1], P. Eisert[1], D. Tatzelt[1], C. Hellge[1], K. Müller[1], T. Ebner[2], S. Bliedung[2]

[1]Fraunhofer Heinrich Hertz Institute, Berlin, Germany, [2]Volucap GmbH, Potsdam-Babelsberg, Germany

ABSTRACT

In June 2018, Fraunhofer HHI together with Studio Babelsberg, ARRI, UFA, and Interlake founded the joint venture Volucap GmbH and opened a commercial volumetric video studio on the film campus of Potsdam Babelsberg. After a testing phase, commercial productions started in November 2018. The core technology for volumetric video production is 3D Human Body Reconstruction (3DHBR), developed by Fraunhofer HHI. This technology captures real persons with our novel volumetric capture system and creates naturally moving dynamic 3D models, which can then be observed from arbitrary viewpoints in a virtual or augmented environment. Thanks to a large number of test productions and new requirements from customers, several lessons have been learnt during the first year of commercial activity. The processing workflow for capture and production of volumetric video has been continuously evolved and novel processing modules and modifications in the workflow have been introduced.

INTRODUCTION

In this paper, some recent developments of the professional volumetric video production workflow are presented. After an overview description of the capture system and the production workflow, several enhancements of the workflow are presented resulting from the experiences gathered after one year of commercial production.

Capture system overview

The capture system consists of an integrated multi-camera and lighting system for full 360 degree acquisition. A cylindrical studio has been set up with a diameter of 6m and height of 4m (see Figure 1 left and right). It is equipped with 32 20MPixel cameras arranged in 16 stereo pairs. The system completely relies on a vision-based stereo approach for multi-view 3D reconstruction and does not require separate 3D sensors. 220 ARRI SkyPanels are mounted behind a diffusing tissue to allow for arbitrarily lit background and different lighting scenarios. This combination of integrated lighting and background is unique. All other currently existing volumetric video studios rely on green screen and directed light from discrete directions, such as the Mixed Reality Studio by Microsoft [1] and the studio by 8i [2].

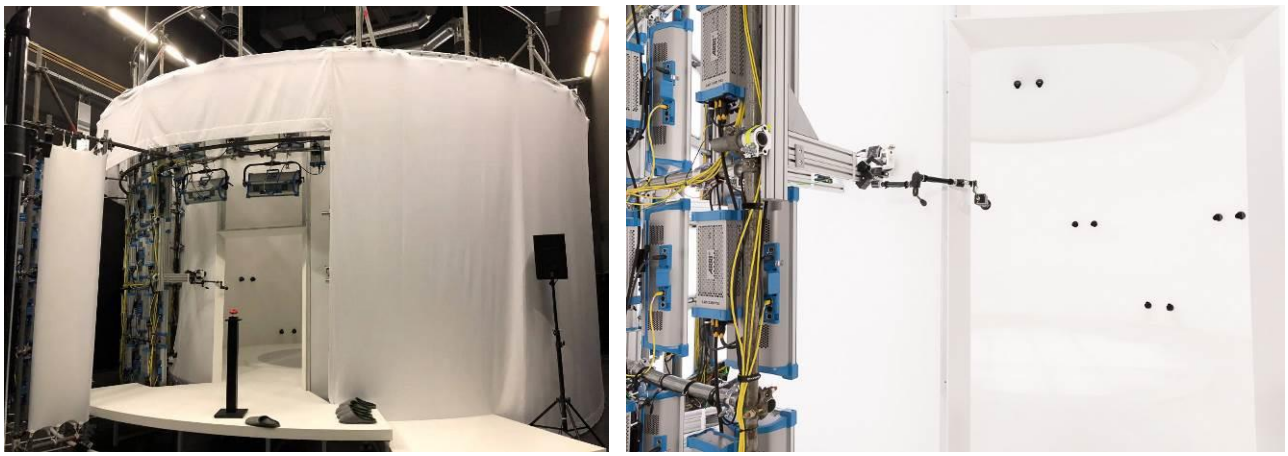

Figure 1 – View from outside the rotunda (left) and inside the rotunda (right)

Workflow overview

In order to allow for professional productions, a complete and automated processing workflow has been developed, which is depicted in the flow diagram in Figure 2. At first, a colour correction and adaptation of all cameras is performed providing equal images among the whole multi-view camera system. After that, a difference keying is performed on the foreground object to minimize further processing. All cameras are arranged in stereo pairs equally distributed in the cylindrical setup. Thus, an easier extraction of 3D information from the stereo base system along the viewing direction is achieved. For stereoscopic view matching, the IPSweep algorithm [3][4] is applied. This stereo processing approach consists of an iterative algorithmic structure that compares projections of 3D patches from left to right image using point transfer via homographic mapping. The resulting depth information for each stereo pair is fused into a common 3D point cloud per frame. Then, mesh post-processing is applied to transform the data to a common CGI (computer generated imagery) format. As the resulting mesh per frame is still too complex, a mesh reduction is performed that considers the capabilities of the target device and adapts to sensitive regions (e.g. face) of the model. For desktop applications, meshes with 60k faces are used, while for mobile devices 20k faces are appropriate. After this processing step, a sequence of meshes and related texture files are available, where each frame consists of an individual mesh with its own topology. This has some drawbacks concerning the temporal stability and the properties of the related texture

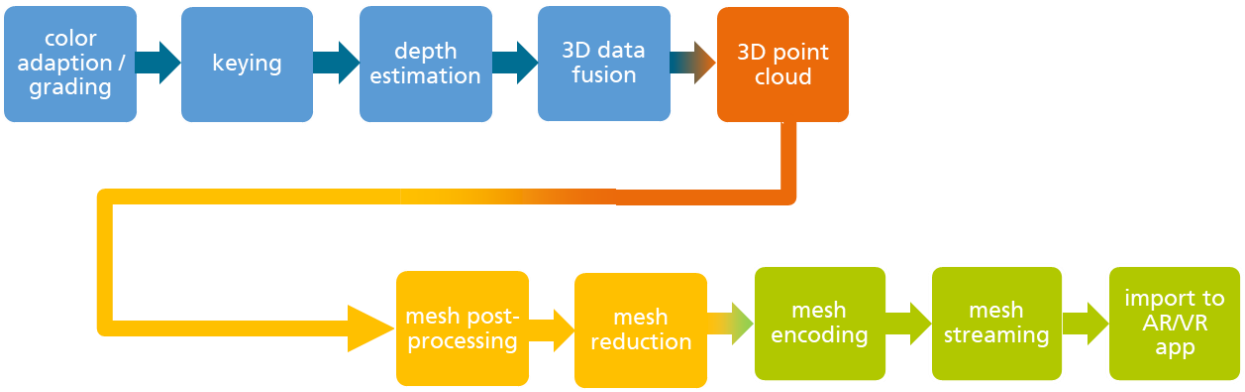

Figure 2 – Production workflow

files. Therefore, a mesh registration is applied that provides short sequences of registered meshes of the same topology. In order to allow the user simple integration of volumetric video assets into the final AR or VR application, a novel mesh encoding scheme has been developed. This scheme encodes the mesh, video and audio independently by using state-of the-art encoding, and multiplexes all tracks into a single MP4 file. On the application side, the related plugin is available for Unity and Unreal to process the MP4 file, decode the elementary streams and render volumetric asset in real-time. The main advantage is a highly compressed bit stream that can be directly streamed from hard disk or via the network using e.g. HTTP adaptive streaming. Unity [5] and Unreal Engine [6] are the two most popular real-time render engines. They provide a complete 3D scene development environment and a real-time renderer that supports most of the available AR and VR headsets as well as operating systems.

PROCESSING-DEPENDEANT COLOUR GRADING

Thanks to the large number of test productions and new requirements from customers, several lessons have been learnt during the first year of commercial activity. The processing workflow for capture and production of volumetric video has been continuously evolved and novel processing modules and modifications in the workflow have been introduced. The unique lighting system offers diffuse lighting from any direction. As a result, the texture of the objects is quite flat without any internal shadows. In Figure 3, left, the original RAW image after capture is shown. This diffuse lighting offers the best possible conditions for re-lighting the dynamic 3D models afterwards, at the design stage of the VR experience. However, the raw image data needs to be adapted according to the needs of different processing steps in the overall volumetric video workflow. Therefore, different colour gradings are applied to support best input data for the following algorithm modules:

- Keying, i.e. segmentation of foreground object from background
- Depth processing
- Creative grading for texturing

For keying, a grading with high saturation is applied in order to optimally distinguish the foreground object from the lit white background. In Figure 3, right, an example of highly saturated grading for keying is depicted. In Figure 4, left, the grading for depth processing is shown. The aim of this grading is to achieve the best possible representation of structures. Hence, especially dark image parts of the person are graded brighter. The diffuse lighting from all directions leads to a very flat image. Therefore and finally, a natural look to the person requires a third way of grading, to recreate a natural skin tone (see Figure 4, right). This final grading is then used for back-projection of the texture onto the final 3D model.

NOVEL MODULES FOR COMPLETION OF END-TO-END WORKFLOW

The two most popular render engines, Unity and Unreal, do not provide optimal tools or workflows for integration of volumetric video. Currently, they can handle animated objects well, but temporarily changing meshes are still challenging. In order to optimally

Figure 3 – Raw image (left), grading for keying (right)

Figure 4 – Grading for depth processing (left), creative grading for texturing (right)

support customers and end users in using this new media format, several additional tools and processing modules were developed. The result of the 3D video processing workflow is independent meshes per frame that consist of individual topology and texture atlas. To create sequences of meshes with identical topology and to improve temporal stability of the related texture atlas, a mesh registration is performed. After definition or automatic selection of a key mesh, succeeding meshes are computed by reshaping the key frame to the geometry of neighbouring frames, while preserving topology and local structures. Bi-directional processing is performed to better deal with sudden topology changes of the mesh sequence. Once the deviations of the registered mesh reach a defined threshold, a new key mesh is set. This mesh registration has two advantages. A new mesh has only to be created for key frames, while for the registered meshes, only 3D vertex positions have to be adapted. Secondly, the texture atlas remains the same due to the same topology of the mesh. A texture atlas example can be seen in Figure 8, left. For this, texture information of adjacent triangles is combined into a texture patch, if the normal vectors

across all triangles do not exceed a given threshold. Otherwise, a new texture patch is created. Then, all texture patches are arranged in a texture atlas, starting with the largest patch in the bottom-left corner. In order to efficiently apply a classical block-based video coding method, two mechanisms are incorporated into the texture atlas: First the topology is preserved across each registered mesh sequence, i.e. shape and position of all patches within the texture atlas remain the same, while only the video content of each patch may change. And second, the empty space between all patches is interpolated (as shown in Figure 8, right), such that the block-based video processing does not produce high rates, when coding image blocks across patch boundaries. In addition to the higher encoding efficiency, this texture atlas processing also allows easier grading of the texture for the registered period of meshes.

The registered mesh sequence is then compressed and multiplexed into an MP4 file and can therefore be easily integrated into dedicated plugins for Unity or Unreal. The meshes are encoded with a standard mesh encoder (see Figure 5). Currently, Draco [7] is used, in which the connectivity coding is based on Edgebreaker [8]. Other mesh codecs are Corto [9], Open 3D Graphics Compression of MPEG and Khronos Group [10], or MPEG activities on Point Cloud Compression [11] (where triangular mesh compression support is currently investigated). The texture atlas is encoded with H.264/AVC due to faster decoding on mobile devices, while an extension to H.265/HEVC is foreseen in the future. This will lead to additional data rate reduction for the compressed stream by keeping the same level of quality. Finally, the audio signal is encoded with a standard audio encoder. The three different elementary streams (ES) are multiplexed into a common MP4 file ready for transmission. On the receiver side, plugins for Unity and Unreal allow for easy integration of volumetric video assets into the target AR or VR application. These plugins include the de-multiplexer, as well as related decoders and perform real-time decoding of the mesh sequence. Typical data rates in our first experiments are 20 Mbit/s[1] for registered mesh sequences targeting mobile devices with 20k faces per mesh. For desktop applications capable of rendering 60k meshes, a data rate for registered meshes with a data rate of 39Mbit/s[2] is achieved. In both experiments, the texture resolution is 2048 x 2048 at a frame rate of 25fps.

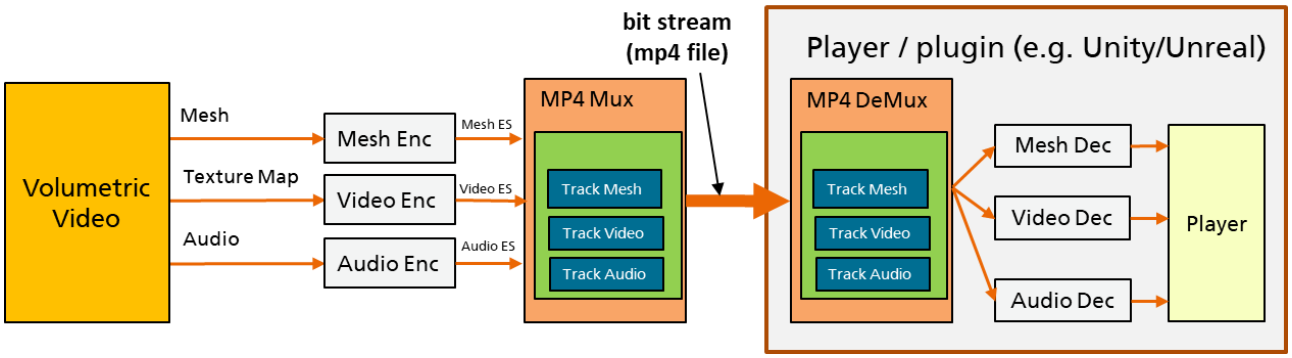

Figure 5 – Mesh encoding, multiplexing, streaming and decoding

[1] 20k Seq.: Individual rates for mesh, video and audio data are 10Mbit/s, 10Mbit/s and 133kbit/s respectively.

[2] 60k Seq.: Individual rates for mesh, video and audio data are 29Mbit/s, 10Mbit/s and 133kbit/s respectively.

PRODUCTION CHALLENGES

Fast movements

The recording of very fast movements, even of individual objects, requires a quick adaptation of the light conditions on the set. Normally, this would only be possible with conversion measures and considerable time expenses during production. The Volucap, however, uses a fully programmable lighting system, which allows light moods or templates to be loaded and used at the touch of a button. This is particularly important as it permits easy reaction to changing requirements during production. In the example with the basketball player Josh Mayo (see Figure 6), special light adjustments for recording very fast movements could be offered by the flexible light control without losing time for the mounting of lights.

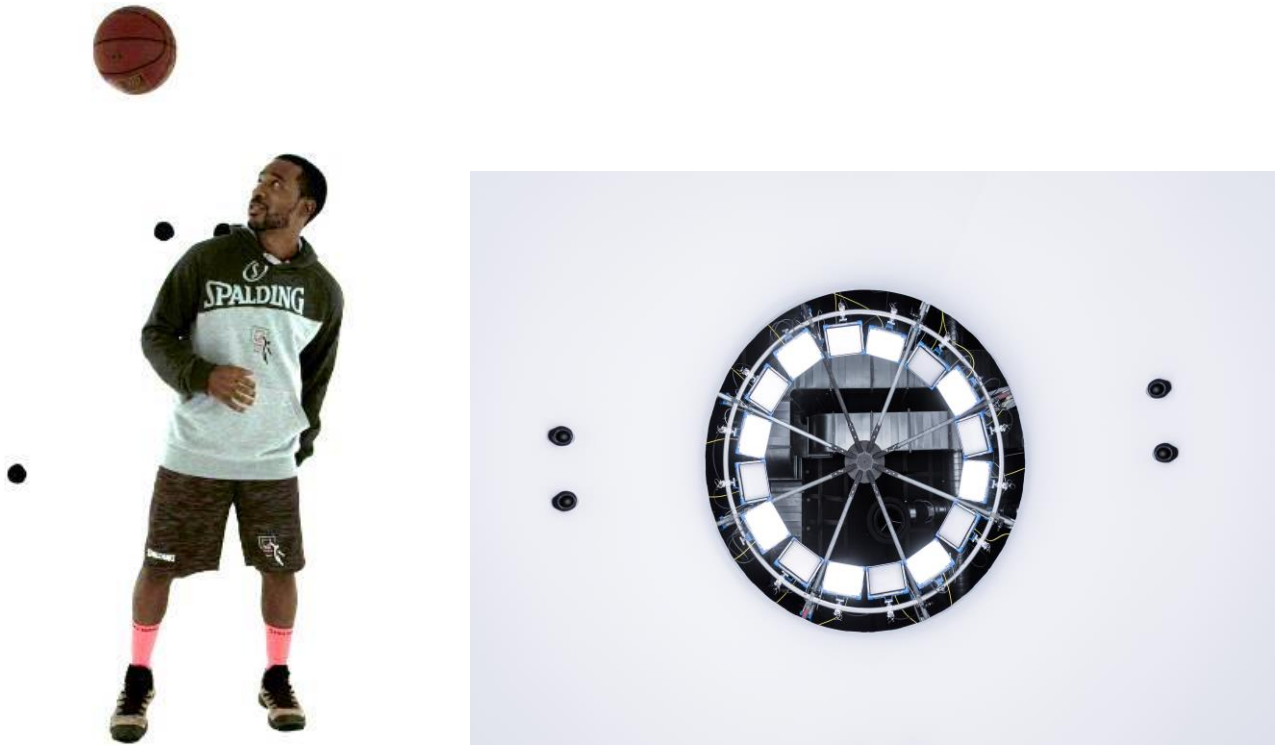

Figure 6 – Fast movements like those of a basketball player (left) require high standards of the lighting system: view into the ceiling of the capture stage (right)

Relighting of actors

For a convincing integration of an actor into a virtual scene it is necessary to have flexibility to adjust the lighting afterwards (see Figure 7). For example, if the exact final illumination in the 3D scene was not known at the time of shooting, or if certain light settings were necessary due to special clothing or movements.

Convenient ways have therefore been developed to transfer the certain lighting of the 3D environment to the actors. These integrate seamlessly with current pipelines such as Autodesk Maya, Nuke or Houdini. A before-after comparison of an original texture atlas and the same texture with integrated ambient lighting is shown in Figure 8.

Figure 7 – Same actor with different lighting

Figure 8 – Original texture atlas (left), texture with integrated ambient lighting (right)

Treadmill

The cylindrical studio offers a recording volume with an averaging diameter of 3m, which is quite comfortable for many productions. However, it is not possible to record more extensive movements, for example walking along a road, without modification. This spatial limitation is by-passed with the help of a treadmill (see Figure 9). This allows two different approaches to be taken. On the one hand, a walking movement can simply be recorded over a longer period of time. On the other hand, a short walking cycle can be recorded and played back in a loop later, as is also the case with conventional animations.

SUMMARY

After one year of commercial production in the new volumetric video studio of Volucap GmbH, a number of modifications and improvements in the workflow have been performed A major improvement related to processing is the introduction of different gradings of the original footage to accommodate the requirements of the workflow. In order to provide a complete end-to-end workflow satisfying the needs of the customers, a mesh registration and mesh encoding scheme has been developed. This significantly reduces the overall amount of data of the volumetric assets and allows for easy integration into standard render engines. Finally, specific challenges have been tackled that resulted from recent

Figure 9 – Extensive walking movements captured on a treadmill

productions. Hence, examples were given for fast moving actors, creative relighting according to the final VR scene and continuous walking scenarios. The huge demands on this new capture and production technology will lead to additional new challenges, which will be researched and solutions developed by Fraunhofer HHI and Volucap together.

REFERENCES

[1] https://www.microsoft.com/en-us/mixed-reality/capture-studios
[2] https://8i.com/.
[3] W. Waizenegger et al., “Real-time Patch Sweeping for High-Quality Depth Estimation in 3D Videoconferencing Applications“ SPIE Conf. on Real-Time Image and Video Processing, San Francisco, USA, (2011). DOI: 10.1117/12.872868
[4] W. Waizenegger, I. Feldmann, O. Schreer, P. Kauff, P. Eisert: Real-time 3D Body Reconstruction for Immersive TV, Proc. 23rd Int. Conf. on Image Processing (ICIP 2016), Phoenix, Arizona, USA, September 25-28, 2016.
[5] https://unity.com
[6] https://www.unrealengine.com
[7] Google. Introducing Draco: compression for 3D graphics. https://opensource.googleblog.com/2017/01/introducing-draco-compression-for-3d.html - last visited: 20th June 2019.
[8] J. Rossignac, “Edgebreaker: Connectivity compression for triangle meshes,” IEEE Trans. Visualization Comput. Graphics, vol. 5, no. 1, pp. 47–61, 1999.
[9] Corto: https://github.com/cnr-isti-vclab/corto - last visited: 20th June 2019.
[10] https://github.com/KhronosGroup/glTF/wiki/Open-3D-Graphics-Compression - last visited: 20th June 2019.
[11] S. Schwarz, et al.: “Emerging MPEG Standards for Point Cloud Compression”, IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 9, issue 1, March 2019