New Experiences of Art via Volumetric Video and Virtual Reality

Dr. Ralf Schäfer[a] and Georg Tschurtschenthaler[b]

[a] Fraunhofer Heinrich Hertz Institute, Germany, ralf.schaefer@hhi.fraunhofer.de; [b] Gebrüder Beetz Filmproduktion Berlin GmbH & Co. KG, Berlin, g.tschurtschenthaler@gebrueder-beetz.de

ABSTRACT: "The Master's Vision" is a three-part VR series that explores the hidden stories behind three outstanding masterpieces. Each episode focuses on one masterpiece and narrates an unknown story about the artwork, offering a new perspective on the masterpiece, its creation or the artist's intention. While VR has already established itself in many areas mainly in the entertainment sector, there are just a few outstanding cultural projects, in which the potential of the technology is exploited and used for a new experience of art. This paper describes the key technology 'volumetric video', which has been used to create parts of the immersive cultural experience and the production of one of the above mentioned episodes.

1. INTRODUCTION

"The Masters Vision" is created in a broad, cross-industry network of independent production companies (gebrueder beetz filmproduktion), TV stations (Arte), research institutions (Fraunhofer HHI) and in close collaboration with three internationally renowned museums (Musée d'Orsay, Alte Nationalgalerie and Munch Museum). One goal of the project is to explore new ways of development and production, and thus find innovative models of co-operation to convey art and other cultural contents. Each episode points out one outstanding work of art, one emotion and an untold story about the masterpiece, which we experience emotionally through the use of Virtual Reality (VR).

Thanks to the availability of new head mounted displays (HMD) for VR, the creation of fully immersive environments has gained a tremendous push. In addition, new Augmented Reality glasses and mobile devices reach the market that allow for novel mixed reality experiences. Also mobile devices are now are capable to register their environment and put CGI objects at fixed positions in viewing space. There are many application in different fields for these technologies, such as gaming, entertainment, virtual prototyping, planning, architecture, construction, chemistry, environmental studies, energy, e-learning and edutainment. In addition there is also a high potential for the cultural sector.

For many of these applications a realistic and lively representation of human beings is desired. However, current character animation techniques do not offer the necessary level of realism. The motion capture process is time consuming and cannot represent all detailed motions of an actor, especially facial expressions and the motion of clothing. This can be achieved with a new technology called Volumetric Video. The main idea is to capture an actor with multiple cameras from all directions and to create a dynamic 3D model of it.

This technology was used for a VR production telling a story about the masterpiece "The Monk by the Sea" by Caspar David Friedrich. This masterpiece is considered as one of the most important paintings of German art history and is a major exhibition piece in the "Alte Nationalgalerie" in Berlin. To translate the central figure of the monk into a VR experience, a volumetric video has been recorded at Fraunhofer Heinrich Hertz Institute, which was then been integrated as a 3D model into a VR experience, in which the user can explore the painting.

In section 2 the artistic background and the story behind the production described in this paper are presented. In Section 3 the

volumetric capture system is described with its main feature of a combined capturing and lighting approach. In section 4, the underlying multi-view video processing workflow is presented. Section 5 provides some results of the recent production and finally the paper concludes with a summary.

2. VR SERIES "THE MASTERS VISION"

Just as feelings guide the creation of every piece of art, feelings also determine the perception of the viewer. Emotions create the connection between the viewer and the artist's vision. The three-part VR series "The Master's Vision" focuses on the emotional understanding of some of the most important masterpieces in European art. In virtual reality emotions can be transported far better than in traditional documentary films. In VR, the user is no longer a passive viewer, but he is himself part of a fantastic, self-contained world, which he can discover.

2.1 THE THREE PAINTINGS

In the series "The Master's Vision" the viewer explores three masterpieces from three different museums. Each episode focuses on a masterpiece and tells a special story about the painting, its creation or the artist's intention.
Based on the main concept of the series, the VR experience starts in an abstract room with excerpts of letters of Caspar David Friedrich. Step by step the viewer will experience the Romantic landscape and the Romantic desire for a metaphysical experience. We can almost feel the sea spraying at our faces, and the wind in our back. The desire to be in a parallel world and the mystification of nature, which is so typical for the Romantic period, is part of the VR experience.

Figure 1: *Caspar David Friedrich "Monk by the Sea"*

We use abstract colours and 3D sounds to let the viewer feel the waves approaching us, to feel that the sky is covered with clouds and that nature is mighty and the human being only a little element.

- *Water Lilies by Claude Monet, Musée d'Orsay Paris (FR)*

Figure 2: *Claude Monet "Water Lilies"*

In 1883 the painter Claude Monet transformed an existing small pond at his house in the French town of Giverny into a water garden with water lilies and a Japanese-style bridge from which he could observe the water and the flowers. From 1899 to 1926, Monet painted more than 250 scenes devoted to the water lily theme, which became what he himself called "an obsession." He has spent more than 30 years of his life trying to paint the reflection of the light on the pond.

Through a contemplative VR experience, that will bring the user from the garden to the Orangerie Museum, passing by the painter's studio, we will give the viewer the opportunity to relive an ongoing beginning of nature, to explore time and space around the Water lilies.

- *The Sun by Edvard Munch, Munch Museum Oslo (NOR)*

'The Sun' is a force of nature. It is wild, energetic, and overpowering in its brightness. With this work, Edvard Munch, known for his dark and gloomy paintings, paints a fiery ode to the bringer of light and warmth after the long Norwegian winter. Munch loved it as an ancient symbol of life, but was also interested in the power of light. What does light do to our brain? Can we see beyond the physical boundaries of our eyes? Can we see beyond the physical boundaries of our eyes?

Figure 3: *Edvard Munch "The Sun"*

By using VR to lead the viewer into this empty space, we extend the experiment on the effect of colour and light on the retina that Munch started in the original set of paintings. Our piece references the work of contemporary light artists like James Turrell, who masters the technique of the ganzfeld effect - the phenomenon of perception without stimulation. The isolating effect of VR headset and earphones creates the perfect circumstance for ganzfeld-induced visions - seeing beyond the physical boundaries of the eye, like Munch wished for a century ago.

Visually, each episode has its own artistic appeal, based on the mood of the artwork. In addition, every piece of art in the series focuses on sensual experience. Which feeling underlies the artwork - and how can the user empathize with it? The style of the artist is not copied or reconstructed - rather, each episode translates the message of the work of art into its own, modern imagery. The VR series goes beyond traditional 360° videos, mixing 3D animations, volumetric videos and user experience in an innovative way.

"The Master's Vision" provides a new understanding of some of the world's most loved works of art. Because Virtual Reality not only allows the viewer to understand, but to feel what makes the artwork so special. As soon as he enters the virtual world, an emotional journey begins, which brings the viewer closer to the intention of the master and lets him experience what the artist wanted to express with his work. Through this kind of mediation, the VR series stands out from audio guides in museums or classic TV production and creates significant added value for the viewer.

2.2 EPISODE "MONK BY THE SEA"

Regarding the cooperation with the Fraunhofer Heinrich Hertz Institute, the episode "Monk by the Sea" deserves a special mention. The aim of this VR episode is to give the viewer the monk's feeling of loneliness and thereby bring him closer to the state of mind of Caspar David Friedrich.

In 2015 the masterpiece by Caspar David Friedrich was restored and it turned out he had originally planned the painting differently. Through special x-rays, a "picture under the picture" became visible. Sea gulls had sketched over the water and ships sailing on the waves. For two years, Friedrich worked intensely on the painting and finally decided against this classic image construction.

The experience starts with personal records and reflections of Caspar David Friedrich. With precise words he describes the setting of his painting. The experience is located on the coast, its surroundings are drawn in bold lines.

We encounter the lonely monk on the beach. While in the painting we only see his back, the viewer can walk around him and get a view of his face. The encounter with the volumetric video of the monk is the heart of the experience. But the monk stays a mysterious character. While the viewer turns away the monk disappears, and the viewer stays alone on the beach in black and white.

A subtle interaction helps the viewer tracing the masterpiece: towards the user looks, one of the seagulls follows him by chance. Directed by an algorithm the flight path of the seagull is like a brushstroke at the same time: the viewer uses his gaze to control the transformation of the image and gradually adds layers of paint until the scenery shines in color.

Figure 4: *Capturing of "the Monk"*

Center of the artwork "Monk by the Sea" and our VR experience is the mystic figure of the monk. Art historians interpret the figure of the

monk often as a self-portrait of the painter. We generated a 3D model of an actor dressed like the monk in Caspar David Friedrich's painting (see Fig. 4). To make sure the users can approach very close to the figure and engage with the character, we had to achieve that our monk is as detailed as possible. The user should recognize and observe the monk´s body language, mimic and gestures. Accepting the figure of the monk as an authentic person was the main principle for an emotional engagement. Instead of animating and rigging the monk we favored a human body reconstruction at Fraunhofer Heinrich Hertz Institute volumetric Studio. The following lines will elaborate the process of 3D volumetric video.

3. VOLUMETRIC VIDEO STUDIO

A novel integrated multi-camera and lighting system for full 360-degree acquisition of persons has been developed at Fraunhofer HHI [1]. It consists of a metal truss system forming a cylinder of 6m diameter and 4m height. In this system 32 cameras are arranged as 16 stereo pairs and equally distributed at the cylindrical plane in order to capture full 360 degree volumetric video. In Fig.2, the construction drawing of the volumetric studio is presented. For illumination 120 LED panels are mounted outside of the truss system and a semi-transparent tissue is covering the inside to provide fully diffuse lighting from any direction. The avoidance of green screen and provision of diffuse lighting from all directions offers best possible conditions for re-lighting of the dynamic 3D models afterwards at design stage of the virtual reality experience. This combination of integrated lighting and background is unique.

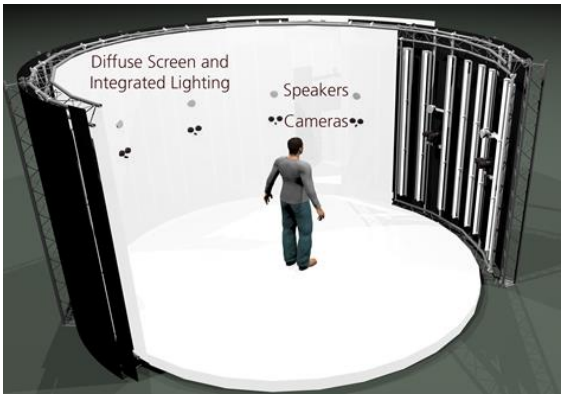

Figure 5: *Scetch of Volumetric Studio*

The system relies completely on a vision-based stereo approach for multi-view 3D reconstruction and omits separate 3D sensors. The cameras are equipped with 20 Mpixel sensors providing a 5k x 4k resolution at 30 frames per second. This is another key difference compared to other existing volumetric video capture systems as this approach benefits from experience in photogrammetry, where high quality 3D reconstruction can be achieved using ultra-high resolution images. The overall ultra-high resolution video information from all cameras lead to a challenging amount of data, resulting in 1.6 TB per minute. In Fig. 3, a view inside the rotunda is shown, with an actor sitting in the center.

Figure 6: *View inside the rotunda during the first test production*

An important aspect is the number and distribution of cameras. The objective was to find the best possible camera arrangement with the least possible number of cameras, whereas, at the same time, the largest possible capture volume with minimum amount of occlusions had to be achieved. In Fig. 7, a sample view of all the 32 cameras is presented that represents our solution for the multi-dimensional optimization problem.

Figure 7: *Single views from the 32 cameras of the volumetric studio*

Meanwhile the development has been so much advanced, that a commercial exploitation

became possible. Fraunhofer HHI founded together with Studio Babelsberg AG, Interlake GmbH, ARRI AG and UFA GmbH the Volucap GmbH at the studio complex in Potsdam Babelsberg (see Fig. 8).

Figure 8: *Albrecht Gerber, former Minister for Economical Affairs and Energy in Brandenburg, Sven Bliedung, CEO Volucap GmbH, and Emilia Schüle, actress, opened the studio in Babelsberg in June 2018 (left); Look into the rotunda with mounted cameras (right)*

This studio is slightly larger than the one at HHI, has a more sophisticated lighting system (over 200 programmable LED panels) and more powerful recording and processing systems. Up to 1.5 hours can be recorded directly onto discs, which is equivalent to a storage capacity of 150 TB.

4. PROCESSING AND DATA FORMATS

In this section, the complete workflow for the processing of volumetric video is described and shown in the workflow diagram in Fig. 9a and 9b. In the first step, the single videos are captured and preprocessed. This prepocessing consists of a color matching to guarantee same color for the same parts of the object in all camera views. This has significant impact on stereo depth estimation, but even more important, it improves the overall texture quality in the point cloud fusion step and the final texturing of the 3D object. In addition, color grading can be applied as well to match the colors of the object with artistic and creative expectations. After color matching and grading, the foreground object is segmented from background in order to reduce the amount of data to be processed. The segmentation approach is a combination of difference and depth keying.

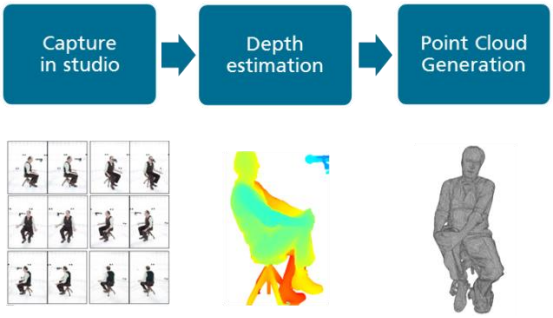

Figure 9a: *Processing chain for volumetric video production (1)*

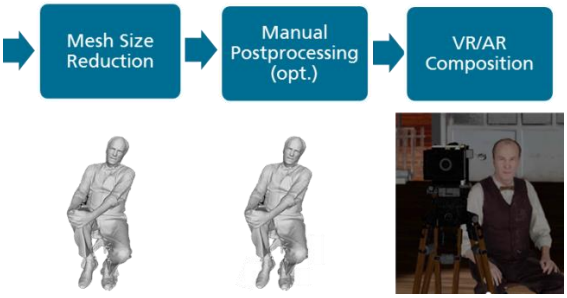

Figure 9b: *Processing chain for volumetric video production (2)*

4.1 STEREO DEPTH ESTIMATION

The next step is stereo depth estimation. As mentioned in the previous section, the cameras are arranged in stereo pairs that are equally distributed in the cylinder. These stereo base systems offer the relevant 3D information from their viewing direction. A stereo video approach is applied that is based on the so-called IPSweep algorithm [2],[3]. In contrast to many other approaches that evaluate a fixed disparity range, a set of spatial candidates and a statistically guided update for comparison is used in this algorithm, which significantly speeds up correspondence search. Moreover, a GPU-centric implementation allows for an inherent sub-pixel processing.

4.2 POINT CLOUD FUSION

As an additional result of stereo processing, initial patches of neighbored 2D points can be calculated straight away including normal information for each 3D point. The resulting 3D information from all stereo pairs is then fused with a visibility-driven patch-group generation algorithm [4]. In brief, all 3D points occluding any other depth maps are filtered out resulting in advanced foreground segmentation. Remaining artefacts have a bigger distance to the object to be reconstructed and as a result they do not occlude any other depth maps. The efficiency

of this approach is given through the application of fusion rules that are based on an optimized visibility driven outlier removal, and the fusion taking place in both, the 2D image domain as well as the 3D point cloud domain. The used algorithms lead to best possible texture with smallest amount of artefacts. Due to the high-resolution original images, the resulting 3D point cloud per frame is in a range of several 10s of millions of 3D points. In order to match with common render engines, the 3D point cloud needs to be converted to a single consistent mesh.

4.3 STEREO DEPTH ESTIMATION

Therefore, a geometry simplification is performed that involves two parts: In a first step, a screened Poisson Surface Reconstruction (SPSR) is applied [5]. SPSR efficiently meshes the oriented points calculated by our patch fusion and initially reduces the geometric complexity to a significant extent. In addition, this step generates a watertight mesh. Holes that remained in the surface after the reconstruction due to complete occlusion or data imperfections are closed. Secondly, the resulting mesh is elementally trimmed and cleaned based on the sampling density values of each vertex obtained by SPSR. In contrast to the common approaches from literature, we do not require an extensive intersection of the resulting surface with the visual hull. Outliers and artifacts are already reliably removed by our patch fusion.

Subsequently, the triangulated surface is simplified even further to a dedicated number of triangles by iterative contraction of edges based on Quadric Error Metrics [6]. Thus, detailed areas of the surface are represented by more triangles than simple regions. During this stage, we ensure the preservation of mesh topology and boundaries in order to improve the quality of the simplified meshes. Another important aspect is the possibility to define the target resolution of meshes. Depending on the target device, a different mesh resolution is necessary in order to match with the rendering and memory capabilities. For a desktop application using Oculus Rift or HTC Vive, a mesh size of 70k faces is appropriate. However, mobile devices such as GooglePixel can render mesh sequences of 20k faces fluently.

The final sequence of meshes can then be further manipulated in standardized post-production workflows, but it can also directly be rendered in virtual reality players, such as Unity3D or Unreal for head mounted displays.

5. PRODUCTION RESULTS

The shooting in the volumetric studio was the main production step in the project. The final production of the piece will be finished at the end of 2018.

- *3D Player*

Unity 3D has been selected as player. The monk's volumetric video, which is the central element of the experience, will be positioned in the VR environment.

Figure 10: *Screen shot of the VR environment*

- *Recording and implementing the inner monologue*

Based on the diaries of Caspar David Friedrich we have recorded a monologue of the artist, that gives us access to the thoughts, and the vision of the painter.

- *Implementation of the interaction*

The interaction between the user and the VR piece is a central element in the experience. The “Monk by the Sea” VR experience can be navigated by the gaze of the user. If the user looks at the monk you hear his voice or inner monologue and you automatically advance towards him – until you stay eyeball to eyeball to him. On a specific point of the experience the monk disappears and the user is alone on the beach. The soundscape changes and a sea gull follows your view direction and starts painting the scape. Like a brushstroke the original masterpiece appears over the sketched landscape until the painting is finished.

- *Sound Design*

The sound of the episode “Monk by the Sea“ tells its own story of loneliness. The view direction of the user changes the sound of the

experience. When the sight wanders across the scenery, a deep sound is audible, which is subtly accompanied by classical instruments and the sound of the sea.

The implementation of the volumetric video was the most complex step during the production. The high quality and the realism of the video had a huge impact on the viewer, and his expectations. Some viewers wanted to interact or to establish eye contact with the monk, others observed him from a distance. Once more we realized how differently people experience VR projects.

During several rounds of testing we optimized the positioning of the monk in the experience and are very curious to learn how a large audience will react on it.

6. SUMMARY

"The Masters Vision" is created as an important pilot project in a broad, cross-industry network of independent production companies (gebrueder beetz filmproduktion, High Road Productions), TV stations (Arte), research institutions (Fraunhofer HHI) and in close collaboration with three internationally renowned museums (Musée d'Orsay, Alte Nationalgalerie and Munch Museum). One goal of the project is to explore new ways of development and production, and thus establish innovative models of co-operation to convey art and other cultural contents.

Volumetric Video is the key technology behind the production of the first episode of the "The Masters Vision" series. This technology uses a large number of cameras creating a data volume of 1.7 TB per minute in order to create natural looking dynamic 3D models of persons. It has meanwhile become so mature, that a professional studio based on HHI's technology and run by Volucap GmbH has been founded on the film studio campus in Potsdam Babelsberg.

The three VR experiences of the series will be marketed and distribute as packages or as stand alone projects. The distribution strategy is based on the following elements:

- arte 360°-App

A 360°-version of the project will be released on the ARTE360 VR app in spring 2019 with the goal to reach a large and international audience.

- Installation in the museums

The most important platform for distribution is the implementation of a 6DoF installation in the museum.

The fact that the VR experience is in proximity of the real piece gives a special charm and appeal to the experience. In the case of the Alte Nationalgalerie the visitor of the museum should first experience the VR piece and then see the real painting. We have scanned the room in order to perfectly integrate the experience in the museum.

The Musée d'Orsay has chosen another option: the VR piece is the central element of an exhibition.

The negotiations with the Munch museum in Oslo are still running. We hope to launch the installation of 'The Sun' in spring 2019, while we are currently getting strong interest from other museums in Norway to exhibit the piece. All the installations can also travel to international festivals for new documentary content, such as Sundance, Sheffield or SXSW, and be installed at VR events and showcases.

Furthermore our distribution partner Camera Lucida from Paris is currently setting up a collection of 6 VR pieces on art for international distribution (included the three pieces of our Master's Vision series). There is concrete interest from galleries in Asia to license the exhibition in 2019.

7. ACKNOWLEDGMENT

The responsible persons for the technical development of the volumetric capturing technologies are Ingo Feldmann and Dr, Oliver Schreer of the Vision and Imaging Technologies Department at Fraunhofer HHI under supervision of Peter Kauff.

8. REFERENCES

[1] P. Kauff, I. Feldmann, O. Schreer, R. Schäfer, G. Tschurtschenthaler: Volumetric Video for „The Masters Vision", in "Kultur und Informatik: Hybrid Systems", Verlag Werner Hülsbusch, ISBN: 978-3-86488-128-2, 2018, pp. 71-86

[2] Waizenegger, Wolfgang 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

[3] Waizenegger, Wolfgang, Feldmann, Ingo, Schreer, Oliver, Kauff, Peter, Eisert, Peter: Real-time 3D Body Reconstruction for Immersive TV, Proc. 23rd Int. Conf. on Image Processing (ICIP 2016), Phoenix, Arizona, USA, September 25-28, 2016.

[4] Ebel, Sascha, Waizenegger, Wolfgang, Reinhardt, Michael, Schreer, Oliver, Feldmann, Ingo: Visibility-driven Patch Group Generation, IEEE Int. Conf. on 3D Imaging (IC3D), Liege, Belgium, December 2014, Best Paper Award.

[5] Kazhdan, M., Hoppe, H.: Screened Poisson Surface Reconstruction, ACM Transactions on Graphics (TOG) 32, No. 3, (2013). DOI: 10.1145/2487228.2487237

[6] M. Garland, m., Heckbert, P. S.: Surface simplification using quadric error metrics, SIGGRAPH '97, Proc. of the 24th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co., New York, USA, 209-216 (1997) DOI: 10.1145/258734.258849