Archiving the Memory of the Holocaust

Ernst Feiler[1], Frank Govaere[1], Philipp Grieß[2], Simon Purk[2], Ralf Schäfer[3](✉), and Oliver Schreer[3]

[1] UFA Serial Drama, Potsdam, Germany
[2] UFA Show & Factual, Potsdam, Germany
[3] Fraunhofer Heinrich Hertz Institute (HHI), Berlin, Germany
ralf.schaefer@hhi.fraunhofer.de

Abstract. Volumetric Video is a rather new technology, which allows the creation of dynamic 3D models of persons, which can then be utilized like computer generated models in any 3D environment. In a recent project between UFA and Fraunhofer HHI a VR documentary about the last German survivor of the Holocaust Ernst Grube has been produced. It consists of six interviews with Ernst Grube lasting about 8–12 min each. The Jewish contemporary witness talks about his experience in Nazi Germany and his imprisonment in the concentration camp Theresienstadt. The VR experience allows the user to meet Ernst Grube and the young interviewer at different places, for which a thrilling virtual environment has been built. Additional interactive components provide the user with more detailed historical information, such as videos, images and text.

Keywords: Volumetric video · Virtual reality · Holocaust documentary

1 Introduction

Thanks to new head mounted displays (HMD) for virtual reality, such as Oculus Rift and HTC Vive, 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. With the ARKit by Apple and ARCore for Android, mobile devices are capable of registering their environment and put CGI objects at fixed positions in viewing space. Besides the entertainment industry, many other application domains have potential for immersive experiences based on virtual and augmented reality. In the industry sector, virtual prototyping, planning, and e-learning benefit significantly from this technology. VR and AR experiences in architecture, construction, chemistry, environmental studies, energy and edutainment offer new applications. Cultural heritage sites, which have been destroyed recently, can be experienced again. Finally yet importantly, therapy and rehabilitation are other important applications. For all these application domains, 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

M. Rauterberg (Ed.): HCII 2020, LNCS 12215, pp. 145–155, 2020.
https://doi.org/10.1007/978-3-030-50267-6_12

of clothes. This can be achieved with Volumetric Video. The main idea is to capture an actor with multiple cameras from all directions and to create a dynamic 3D model.

With the availability of this technology the idea was born, to use Volumetric Video also for recording live persons and to archive their memory for future generations, when they will no longer be alive. This becomes extremely important, if the memory of such persons reflects an important period of history and as it is the case for the Holocaust. Although this dark period of German history is only 75 years ago, it disappears out of the thoughts especially of young people. Therefore, it is important to preserve the memories of contemporary witnesses in such a way, that it can be presented in an appealing way to young audiences.

This reflection was the starting point for the documentary about one of the last German survivors of the Holocaust Ernst Grube named "ERNST GRUBE – THE LEGACY", which will be described in this paper. In Sect. 2 the story board of the production is presented, while in Sect. 3 the technologies for generation of volumetric 3D models are explained. In Sect. 4 the construction of the virtual environment, in which the recorded persons are presented and in Sect. 5 the interactive experience will be described. Finally the results of an evaluation of the production will be presented in Sect. 6 before summarizing the paper and giving an outlook in Sect. 7.

2 The Story Board Volumetric Video Production

The short VR film "ERNST GRUBE – THE LEGACY" consists of six interviews with Ernst Grube lasting about 8–12 min each. These interviews have been carried out by a young person, because the idea is to show these short films in schools, museums and memorials and to especially attract a young public in order to inform them about this dark chapter of German history (Fig. 1).

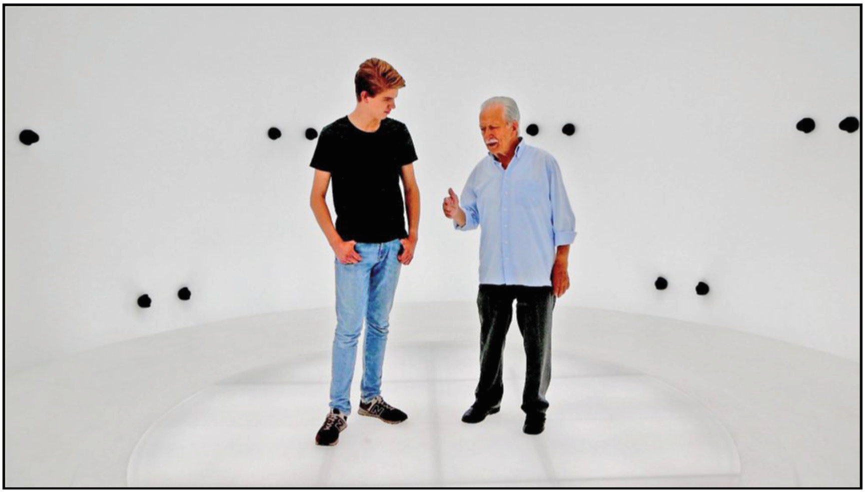

Fig. 1. Student Phil Carstensen with Ernst Grube in the volumetric studio. Photo: UFA

The Jewish eyewitness talks about his experience in Nazi Germany and his imprisonment in the concentration camp Theresienstadt. This "walkable film" represents a time document of compelling authenticity.

Altogether six stations of his life are told, which are: (1) the exclusion of the Jewish population by the Nazi regime, (2) the Jewish life in Nazi Germany, (3) his life in the ghetto in Munich, (4) the fear of deportation, (5) the concentration camp Theresienstadt and (6) his life in Germany after the Second World War.

3 Volumetric Video Production

There are several companies worldwide offering volumetric capture systems, such as Microsoft with its Mixed Reality Capture Studio [1], 8i [2], Uncorporeal Systems [3] and 4D Views [4]. Compared to these approaches, the presented capture and processing system for volumetric video distinguishes in several key aspects, which will be explained in the next sections. Concerning multi-view video-based 3D reconstruction, several research groups work in this area. A complete workflow for volumetric video production based on RGB and depth sensors is presented in [5]. In [6], a spatio-temporal integration is presented for refinement of surface reconstruction. This approach is based on 68 4M pixel Cameras requiring approx. 20 min/frame processing time to achieve a 3M faces mesh. Robertini et al. [7] present an approach focusing on surface detail refinement by maximizing photo-temporal consistency. Vlasic et al. [8] present a dynamic shape capture pipeline using eight 1k cameras and a complex dynamic lighting system that allow for controllable light and acquisition at 240 frames/sec. The high-quality processing requires 65 min/frame and a Graphics Processing Units (GPU) based implementation with reduced quality achieves 15 min/frame processing time.

In Sect. 3.1, the volumetric capture system is presented with its main feature of a combined capturing and lighting approach. In Sect. 3.2, the underlying multi-view video processing workflow is presented.

3.1 Volumetric Capture

A novel integrated multi-camera and lighting system for full 360-degree acquisition of living persons has been developed. It consists of a metal truss system forming a cylinder of 6 m diameter and 4 m height. On this system, 32 cameras are arranged in 16 stereo pairs and equally distributed at the cylindrical plane in order to capture full 360-degree volumetric video. In Fig. 2, left, the construction drawing of the studio is presented. 120 KinoFlo LED panels are mounted outside the truss system and a semi-transparent tissue is covering the inside to provide diffuse lighting from any direction and automatic keying. The avoidance of green screen and provision of diffuse lighting from all directions offers best possible conditions for relighting of the dynamic 3D models afterwards at design stage of the VR experience. This combination of integrated lighting and background is unique. All other currently existing volumetric video studios use green screen and directed light from discrete directions.

The system completely relies on a vision-based stereo approach for multi-view 3D reconstruction and omits separate active 3D sensors. The cameras are equipped with a

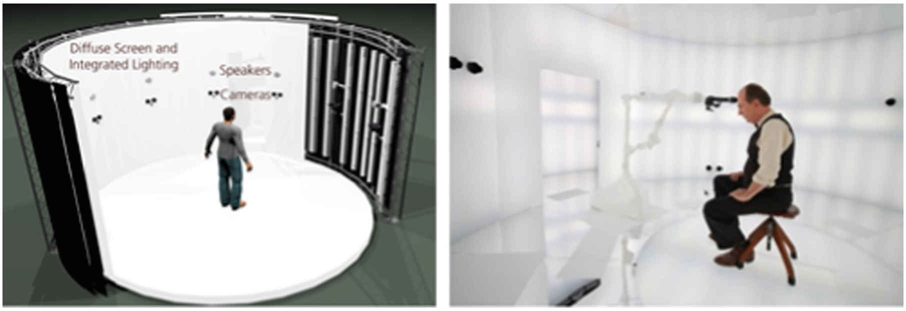

Fig. 2. Drawing of the capture and light stage (left) and first prototype (right)

high-quality 20 MPixel sensor at 30 frames per second. This is another key difference compared to other existing capture systems. The overall ultra-high resolution video information from all cameras leads to a challenging amount of data, resulting in 1.6 TB per minute. In Fig. 2, right, a view inside the rotunda is shown, with an actor sitting in the center.

For the multi-view camera system, the aim was to find the best possible arrangement of least possible number of cameras, with the largest possible capture volume and minimum amount of occlusions. In Fig. 3, a sample view of all 32 cameras is presented that represents our solution for the multi-dimensional optimization problem. Four pairs are mounted on the ceiling and on the bottom, while eight pairs are distributed equally at middle height in the cylinder.

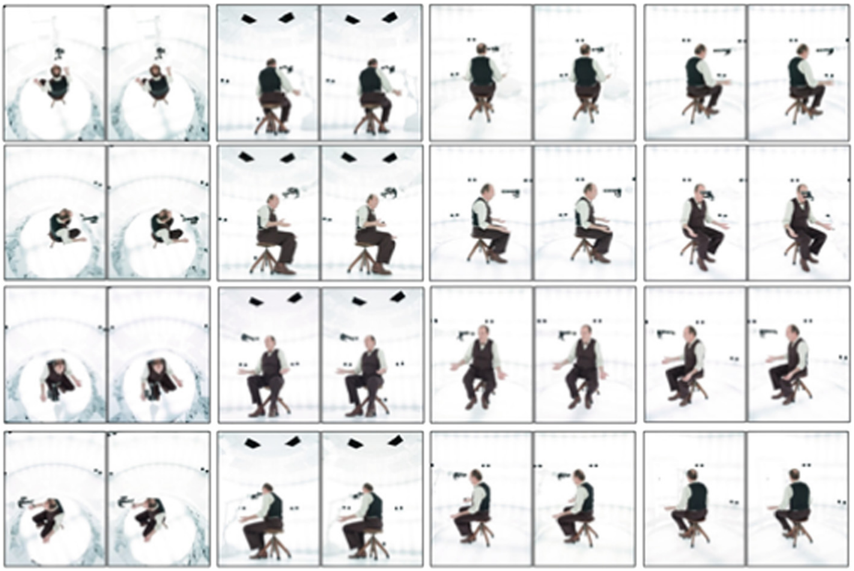

Fig. 3. 32 camera views

3.2 Processing of Volumetric Video

Now, the complete volumetric video workflow is described, consisting of pre-processing, stereo depth estimation, point-cloud fusion, meshing and mesh reduction (see Fig. 4).

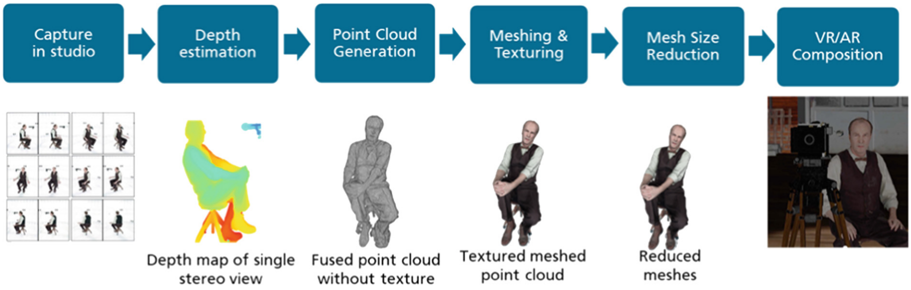

Fig. 4. Production pipeline for volumetric video

Pre-processing. In the first step, a pre-processing of the multi-view input is performed. It consists of a color matching to guarantee consistent colors in all camera views. This has significant impact on stereo depth estimation, but even more important, it improves the overall texture during 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. E.g. colors of shirts can be further manipulated to get a different look. After that, 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 supported by the active background lighting.

Stereo Depth Estimation. The next step is stereo depth estimation. As mentioned before, 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 IPSweep algorithm [9, 10]. 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. Once all candidates are evaluated for a given similarity measure, the best candidate is selected as final depth candidate. Compared to standard block-matching approaches, spatial 3D patches are projected from the left to the right image (as well as from the right to left) in order to consider perspective distortions. After that, a consistency check is performed between both depth maps and a consistency map is produced. This is used to hinder inconsistent matches to propagate in the next iteration and to penalize their selection. The iterative structure of the algorithm allows for propagation of results to their local neighborhood, while keeping pixel independent processing, which enables highly efficient parallel implementation on GPU.

Point Cloud Fusion. For each 2D depth map, initial patches of neighbored 2D points can be calculated straight away including information about the normal on the surface for each 3D point. The resulting 3D information from all stereo pairs is then fused with a visibility-driven patch-group generation algorithm [11]. In brief, all 3D points occluding any other depth map are filtered out resulting in an advanced foreground segmentation. 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. Due to the high-resolution original images, the resulting 3D point cloud per frame is in a range of several 10 s 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.

Meshing and Mesh Reduction. A geometry simplification is performed that involves two parts: In a first step, a screened Poisson Surface Reconstruction (SPSR) is applied [12]. 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 introduced earlier, 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 an appropriate number of triangles by iterative contraction of edges based on Quadric Error Metrics [13]. 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. To recover details lost during simplification, we compute UV coordinates

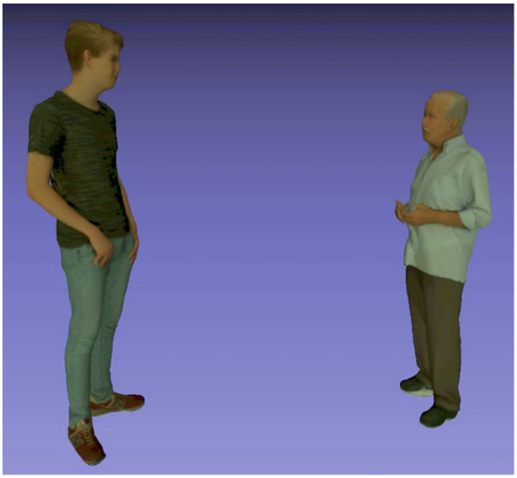

Fig. 5. Result of final volumetric asset

for each vertex and create a texture of suitable size [14]. In Fig. 5, an example of the final volumetric 3D model of both persons is presented.

The final sequence of meshes is then further manipulated in standardized post-production workflows, but also be rendered directly in virtual reality applications, created with 3D engines like Unity3D [15] or Unreal Engine [16].

4 Building the Virtual Environment

The challenge was to recreate the historical sites as locations for the interviews. We believe the eyewitness report will have more impact, when it is told in a convincing reconstruction of the actual historical environment. Some of these locations are no longer in existence and in some cases picture material is scarce. It was therefore necessary to conduct a substantial amount of historical research. As a first step in building the virtual environment concept art was created using these historical references. The tools for this were rather low-tec: pencil and paper (Figs. 6, 7 and 8).

These concept art sketches and the existing photo material were the base for our 3D-artists to start modelling the environments in Autodesks 3DS Max. In order to enable the real time performance, we used low-polygon modelling techniques. Textures were made in Adobe Photoshop, using the Quixel Plugin, which helped us to create so called PBR materials (physically based rendering). This was essential to achieve a photo-realistic result in Unity 3D, a real time engine in which we combine the environment, characters and interactive content (Fig. 9).

5 The Interactive Experience

The final VR experience will allow the user to meet Ernst Grube and the young interviewer at all these different places mentioned in Sect. 2, for which the virtual environment described in Sect. 4 has been built. Additional interactive components provide the user with more detailed historical information, such as videos, images and text.

The life story of Ernst Grube is told in different stations. The various locations are arranged along a path in the form of vignettes (Fig. 10). On the other side of the path are stele-shaped milestones (Fig. 11). These milestones represent the interactive content

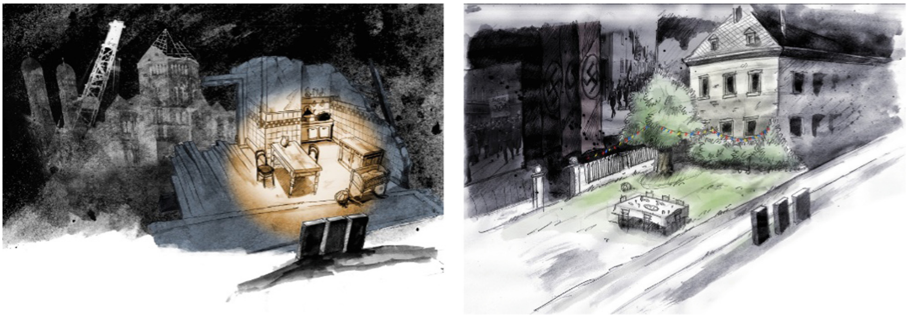

Fig. 6. Concept art of apartment in Munich (left) and children's home (right)

that the user can access. Thus, the set is also the user interface. The user can teleport into the various time segments along the way and control the interactive elements. Archive material is displayed on floating, transparent screens.

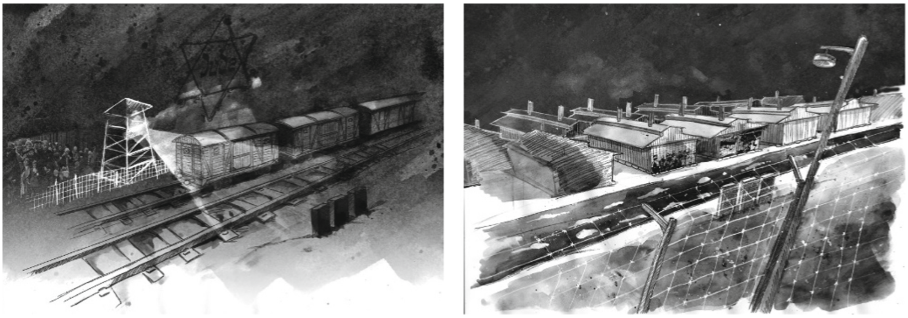

Fig. 7. Concept art of freight yard (left) and deportation camp (right)

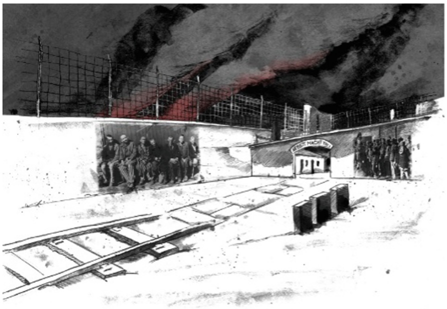

Fig. 8. Concept art of main gate of Theresienstadt

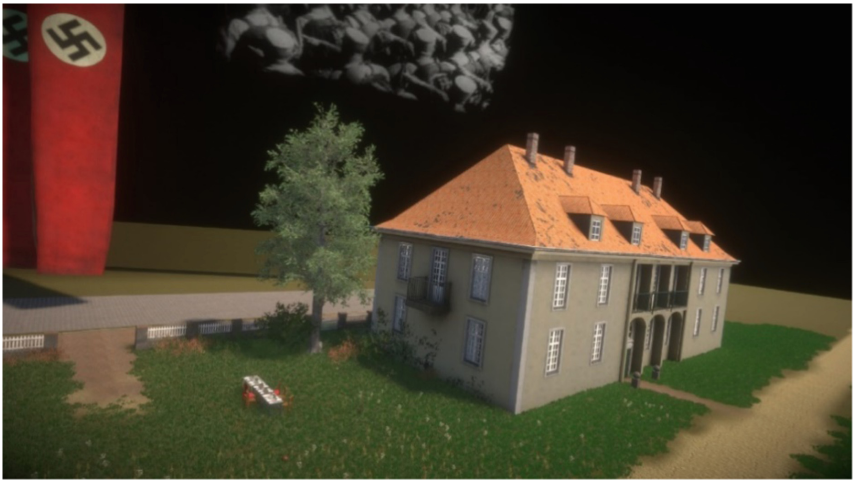

Fig. 9. Screenshot of the VR scene of the children's home

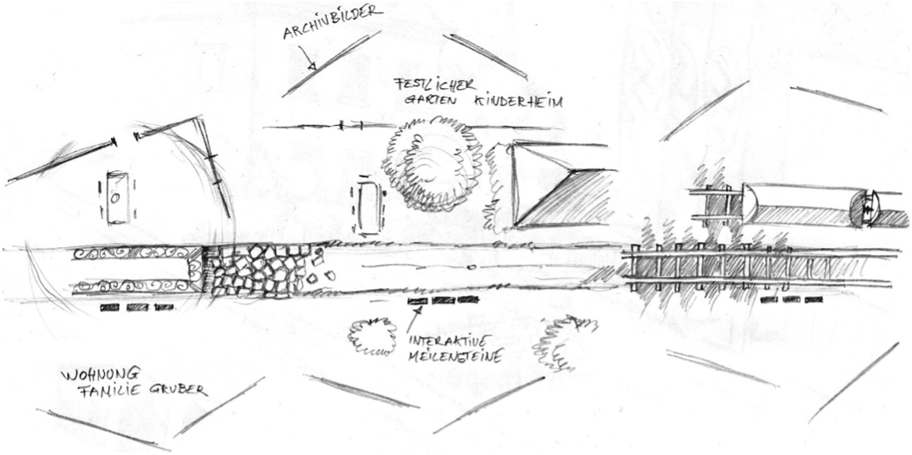

Fig. 10. Interactive set (top view)

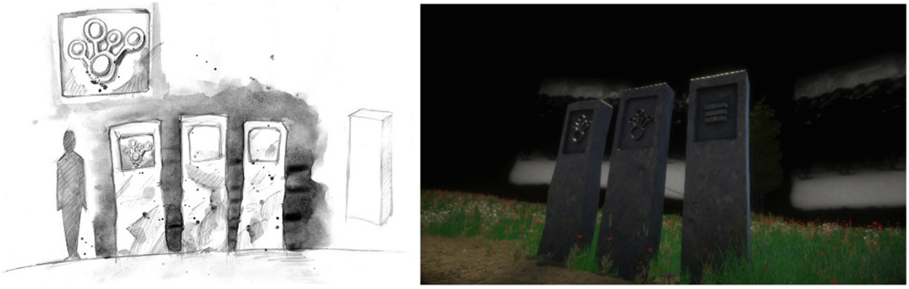

Fig. 11. Drawing of interactive steles (left) and final 3D model (right)

6 Evaluation

One objective of the collaboration between Fraunhofer HHI and UFA GmbH is to evaluate the acceptance and level of experience. Hence, a proof-of-concept VR experience is developed, where the user joins the contemporary witness Ernst Grube and the young student in the garden of the children's home. This proof-of-concept will last for three minutes and will be presented to different audiences and to the public. It will be presented in the visitor center of the memorial site Sachsenhausen, Germany. The VR experience is also demonstrated to executives from the foundation of Brandenburg memorials (Stiftung Brandenburgische Gedenkstätten). One major objective is to use this new concept of interactive storytelling of historical content in secondary schools during history lessons. Therefore, the VR experience will be brought to a Berlin secondary school to let the pupils experience the story of Ernst Grube.

Based on the outcome of the evaluation, the VR experience will be further adopted. With this first proof-of-concept, the general public shall be attracted and convinced that further investment in this new technology is required.

7 Summary and Outlook

In this paper we have presented a system for the production of volumetric video. This system has been used to produce the short VR film "ERNST GRUBE – THE LEGACY", which consists of six interviews with Ernst Grube lasting about 8–12 min each. The idea of this production is to keep the memory of the Holocaust alive by interviewing one of the last German survivors of this dark period in German history and letting him report about different stages of his martyrdom. It is intended to showcase this film at different locations such as the former concentration camp Sachsenhausen, which today is a memorial site. Most importantly, it is planned to use this VR experience in history classes, because it is important to keep young people informed about the felony of the Nazi regime. Therefore, special care has been taken for the design of the experience. Ernst Grube has been interviewed by a young person, which speaks the language of the young generation, authentic historical sites have been constructed as virtual environments and means of interaction have been added, so that young people get excited about the experience.

At this stage only the first of the planned six episodes has been produced, but the other five will follow in the future. Some novel interaction tools are currently under development, which may be added. One option is to establish eye contact between the viewer and the 3D model of Ernst Grube for viewing positions, where the users looks into his face. It is expected, that the feeling of immersion can be further increased by such means.

In addition, UFA expects to perform additional productions with other Holocaust survivors or other individuals of public interest in the future and with it create a new documentary format. As time goes by, this new way of interactive experience of contemporary witnesses plays a significant role to preserve cultural heritage and history of human being.

References

1. https://www.microsoft.com/en-us/mixed-reality/capture-studios
2. https://8i.com/
3. http://uncorporeal.com/
4. 4D View Solutions. http://www.4dviews.com
5. Collet, A., et al.: High-quality streamable free-viewpoint video. ACM Trans. Graph. **34**(4) (2015). https://doi.org/10.1145/2766945. Article 69
6. Leroy, V., Franco, J.-S., Boyer, E.: Multi-view dynamic shape refinement using local temporal integration. In: IEEE International Conference on Computer Vision 2017 (October 2017)
7. Robertini, N., Casas, D., De Aguiar, E., Theobalt, C.: Multi-view performance capture of surface details. Int. J. Comput. Vis. (IJCV) **124**, 96–113 (2017)
8. Vlasic, D., Peers, P., Baran, I., Debevec, P., Popovic, J., Rusinkiewicz, S.: Dynamic shape capture using multi-view photometric stereo. ACM Trans. Graph. **28**(5), 174 (2009)
9. Waizenegger, W., Feldmann, I., Schreer, O.: Real-time patch sweeping for high-quality depth estimation in 3D videoconferencing applications. In: SPIE Conference on Real-Time Image and Video Processing, San Francisco, USA, (2011). https://doi.org/10.1117/12.872868
10. Waizenegger, W., Feldmann, I., Schreer, O., Kauff, P., Eisert, P.: Real-time 3D body reconstruction for immersive TV. In: Proceedings of 23rd International Conference on Image Processing (ICIP 2016), Phoenix, Arizona, USA, September 25–28 (2016)

11. Ebel, S., Waizenegger, W., Reinhardt, M., Schreer, O., Feldmann, I.: Visibility-driven patch group generation. In: IEEE International Conference on 3D Imaging (IC3D), Liege, Belgium, December 2014, Best Paper Award (2014)
12. Kazhdan, M., Hoppe, H.: Screened poisson surface reconstruction. ACM Trans. Graph. (TOG) **32**(3), 1–13 (2013). https://doi.org/10.1145/2487228.2487237
13. Garland, M., Heckbert, P.S.: Surface simplification using quadric error metrics. In: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1997, pp. 209–216. ACM Press/Addison-Wesley Publishing Co., New York (1997) https://doi.org/10.1145/258734.258849
14. Ebner, T., Feldmann, I., Renault, S., Schreer, O.: 46-2: distinguished paper: dynamic real world objects in augmented and virtual reality applications. In: SID Symposium Digest of Technical Papers, Los Angeles, USA, vol. 48, no. 1, pp. 673–676, May 2017, Distinguished Paper Award. https://doi.org/10.1002/sdtp.11726
15. https://www.unity3d.com/
16. https://www.unrealengine.com/