Research Article

Oliver Schreer*, Markus Worchel, Rodrigo Diaz, Sylvain Renault, Wieland Morgenstern, Ingo Feldmann, Marcus Zepp, Anna Hilsmann, and Peter Eisert

Preserving Memories of Contemporary Witnesses Using Volumetric Video

Abstract: Volumetric Video is a novel technology that enables the creation of dynamic 3D models of persons, which can then be integrated in any 3D environment. In contrast to classical character animation, volumetric video is authentic and much more realistic and therefore ideal for the transfer of emotions, facial expressions and gestures, which is highly relevant in the context of preservation of contemporary witnesses and survivors of the Holocaust. Fraunhofer Heinrich-Hertz-Institute (HHI) is working on two projects in this cultural heritage context. 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. A second project started in collaboration with the University Munich, faculty of languages and literature and Geschwister-Scholl-institute for political science, creating a concept for a VR experience together with Dr. Eva Umlauf, the youngest Jewish survivor in the concentration camp in Auschwitz. This paper presents key aspects of volumetric video and details about both projects including a discussion about the user perspective in such a VR experience.

Keywords: Volumetric video, Virtual Reality, interactivity, holocaust, documentary

---

*Corresponding author: Oliver Schreer, Fraunhofer Institute for Telecommunications Heinrich-Hertz-Institute HHI, Vision & Imaging Technologies, Einsteinufer 37, 10587 Berlin,Germany, e-mail: oliver.schreer@hhi.fraunhofer.de, ORCID: https://orcid.org/0000-0001-6900-8287
Markus Worchel, Rodrigo Diaz, Sylvain Renault, Wieland Morgenstern, Ingo Feldmann, Marcus Zepp, Anna Hilsmann, Peter Eisert, Fraunhofer Institute for Telecommunications Heinrich-Hertz-Institute HHI, Vision & Imaging Technologies, Einsteinufer 37, 10587 Berlin,Germany, e-mails: markus.worchel@hhi.fraunhofer.de, RodrigoMauricio.DiazFernandez@hhi.fraunhofer.de, sylvain.renault@hhi.fraunhofer.de, wieland.morgenstern@hhi.fraunhofer.de, ingo.feldmann@hhi.fraunhofer.de, marcus.zepp@hhi.fraunhofer.de, anna.hilsmann@hhi.fraunhofer.de, peter.eisert@hhi.fraunhofer.de

1 Introduction

Archiving memories for future generations becomes an important task as fewer contemporary witnesses of the Holocaust will be alive in the coming years. Although there is huge amount of 2D video material available, there are several attempts to use digital technologies for the creation of immersive interactive media experiences. The "LediZ" project creates holographic testimonies of German Holocaust survivors [1, 2]. A viewer experiences the stereoscopically filmed witnesses using 3D glasses and can vocally query a pool of pre-recorded answers. "New Dimensions in Testimony" [3], a similar project, uses volumetric capture but their display-based presentation can only cover very limited viewpoints.

Thanks to the advances in eXtended Reality (XR) technology, it is possible to create truly immersive experiences with contemporary witnesses. The key challenge is to create highly realistic digital representations of humans. In comparison to computer-generated models, volumetric video has the potential to avoid the so-called uncanny valley effect, where humanoid objects that imperfectly resemble actual human beings provoke uncanny or strangely familiar feelings of eeriness and revulsion in observers [4].

In this paper, we present two projects, where volumetric video technology is used to recreate experiences with the contemporary witnesses Ernst Grube and Eva Umlauf. After a short description of the workflow for creation of volumetric video, the current state of the two projects is presented. Following that, the user experience concepts are discussed as they differ significantly in both projects.

Finally, we present a method of introducing interactive animations to the experience. We have demonstrated the interactivity for other volumetric videos, and are currently working on bringing it to the recently recorded Eva Umlauf project.

2 Volumetric Video Production

Fraunhofer HHI has been working on volumetric video technology for over a decade and built its first system for high-quality dynamic 3D reconstruction in 2011 [5], then in the context of 3D video communication. In 2017, Fraunhofer HHI introduced its first 360 ° studio for volumetric video that was mainly used for test productions with partners like UFA. In 2019, HHI span off the company Volucap GmbH together with UFA, Studio Babelsberg, ARRI and Interlake [6], which is performing commercial productions since the beginning of 2019. In recent years, volumetric video has been of huge commercial interest by other companies [7, 8, 9] and also gained significant traction in the research community (e. g., [10, 11, 12]).

The Volucap studio is based on the Fraunhofer HHI research prototype from 2017. A novel integrated multi-camera and lighting system for full 360-degree acquisition of persons has been developed. It consists of a metal truss system forming a cylinder of 6 m diameter and 4 m height. 120 KinoFlo LED panels are mounted outside the truss system and a semi-transparent tissue covers 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. Within the rotunda, 32 cameras are arranged as 16 optimally distributed stereo pairs. Each camera offers a resolution of 4k × 5k, which results in a data volume of 1.6 TB of raw data per minute. The cameras are fully calibrated, so both their location and sensor characteristics are known. The 16 original left views of each camera pair are depicted in Figure 1. For 3D reconstruction, the captured data is processed by the pipeline depicted in Figure 2. For this work, we use the improved workflow presented in [13].

In the first step, a pre-processing of the multi-view input is performed. It consists of a color matching processing step to guarantee consistent colors in all camera views. This is relevant for stereo depth estimation to support reliable and accurate matching between the two views. Even more important, it improves the overall texture during the final texturing of the 3D object. There, texture information for neighboring surface patches is used from different cameras and equalized colors reduce artifacts. 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 dif-

Figure 1: 16 original left views of each camera pair.

ferent 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. The standard segmentation mode is based on a statistical method [14], which is a combination of difference and depth keying supported by the active background lighting. Recently, an alternative method is used based on a machine-learning model [15], fine-tuned on hundreds of manually labelled images from current and past volumetric capture data. The machine learning segmentation outperforms the statistical per-pixel approach most notably at resolving local ambiguities, i. e., when a foreground pixel has a similar color as the clean plate pixel at the same location. On the other hand, the statistical approach is faster and less heavy on GPU memory.

After that, a stereo depth estimation algorithm based on the Iterative Patch Sweep [5, 16] is used to regress one depth map for each of the 16 camera pairs. The individual depth maps are then back-projected into 3D space and fused into an oriented point cloud, which is cleaned from outliers and wrong predictions.

We join depth maps into a single 3D representation, a truncated signed distance function (TSDF) [17], with an algorithm based on [18]. We first build a discretized 3D volume and project all depths maps into it. Using constraints
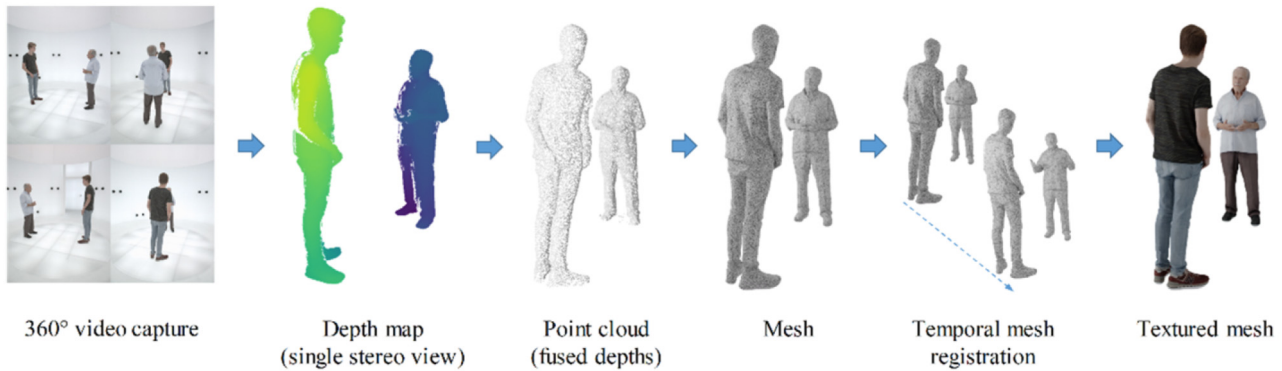
360° video capture   Depth map (single stereo view)   Point cloud (fused depths)   Mesh   Temporal mesh registration   Textured mesh

Figure 2: Volumetric video production pipeline.

from the projected depths, we then compute the distance to potential surfaces for nearby voxels. A point cloud is extracted by finding zero crossings in the grid of this discretized TSDF. Since all depth values are integrated by weighted averaging, Gaussian noise is reduced along the surface. We extend the algorithm above with normal propagation. More precisely, we also average the estimated per-pixel normals. The resulting point cloud normals do not strictly adhere to the surface approximated by the TSDF, but are biased towards the average direction of the points' source cameras. This property is exploited in a later cleaning stage.

Although the TSDF approximates surfaces reasonably well, the resulting point cloud can still contain points not belonging to the protagonists. To tackle this issue, we introduce a two-step method that uses visual hull-like constraints and a "soft" capture volume clipping. First, we project each point onto each camera image plane. Only if a point falls within the image borders and is not covered by the foreground mask, we remove it. Retaining points outside of the frustum is motivated by partial scene coverage (i. e., not every camera sees every part of the studio). In the second step, we divide the scene into multiple sub-volumes depending on the number of cameras covering them. We then remove points from volumes that are only visible in a few cameras.

A triangle mesh is extracted from the point cloud using a modified version of the Screened Poisson Surface reconstruction [19]. This procedure uses the preprocessed masks produced by the segmentation step to visually constrain the geometry silhouette and trim any inaccuracies present due to the iso-surface extraction. After this, the mesh is further cleaned and decimated, reducing the number of triangles from potentially millions to a few ten thousands, depending on the target application. Semantically important regions like the face can thereby be represented by more triangles [20]. With one triangle mesh per frame,

the temporal mesh registration then relates meshes over time to increase the temporal stability of the mesh sequence. A key frame based approach is exploited [21] that tracks the vertices of a reference mesh in forward and backward direction leading to constant mesh topologies. Larger structural changes are handled by the insertion of a new key frame with associated new mesh topology followed by a smooth blending across key frame sequences. Finally, the textured mesh is obtained by projecting all camera images from the corresponding frame onto the surface. Depending on the visibility, triangle size, angle towards each view, and distance from the contours, a weight is computed for each projected pixel in texture space. A texture map is obtained using a weighted mean of the projected colors per texel. Additionally, the resulting map is processed to prevent interpolation artifacts along the mesh seams. The resulting sequence of textured meshes is then encoded with a mesh-encoder for the geometry and a video encoder for the texture. Both streams are multiplexed in an mp4 stream. This mp4-file can then be streamed and rendered via a dedicated plug-in in real time by game engines like Unity or Unreal on devices such as tablets or Augmented Reality (AR) & Virtual Reality (VR) glasses. With our plug-in, the volumetric models can be easily integrated into appropriate virtual scenes.

The major challenge for the development of such a complex multi-view video workflow is to preserve the texture and geometrical detail, while reducing the amount of data significantly. Just to remind, the amount of raw data is 1.6 TeraByte per minute of volumetric video. As shown in the diagram, in Figure 3, the presented workflow reduces the amount of data per one minute of volumetric video by more than three orders of magnitude. The size of the resulting mp4-file is in the order of several hundred Mbyte depending on the chosen level of detail and resolution of the texture. Some examples for the achieved quality are presented in section 3.
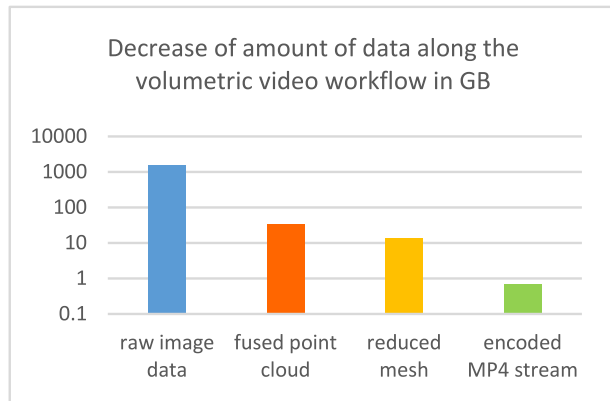
Figure 3: Decrease of data along the pipeline per one minute of volumetric video.

3 Ernst Grube – The Legacy

For many years, Fraunhofer HHI and UFA GmbH have been working together on new media formats and storytelling by using volumetric video. In 2019, the idea was born to create a VR experience, where the user can join a talk between Ernst Grube, one of the last German survivors of the Holocaust and a young student. The story consists of six interviews with Ernst Grube lasting about 8–12 minutes 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 will be built. Additional interactive and visual components provide the user with more detailed historical information, such as videos, images and text.

Figure 4, presents two different camera views of the two protagonists inside the capture stage. Figure 4 – left,

Figure 4: Ernst Grube and the student: two different views, original (left), reconstructed meshes (middle), reconstructed and textured meshes (right).

Figure 5: Integration of the volumetric character in the scene "Children's Home".
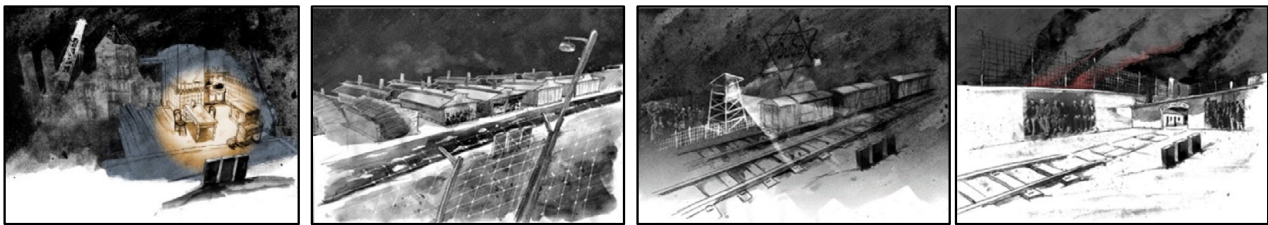

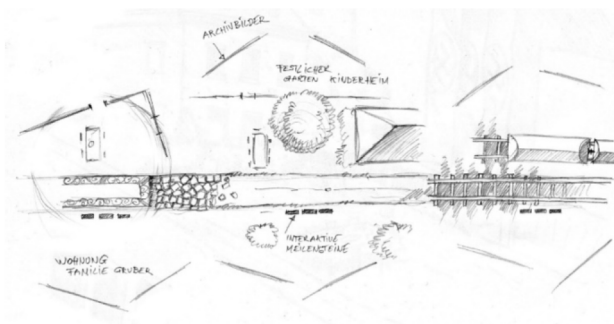**Figure 6:** Conceptual art for other episodes.

Figure 7: Concept of the connected storyline.

shows the original camera view. Figure 4 – middle, shows the related geometry resulting from our volumetric video workflow. Finally, on the right, the textured mesh of the two protagonists is presented from the view of the original camera views in the left.

In Figure 5, the integration of both protagonists in the children's home-episode, the one out of six episodes, is illustrated. The virtual environment is created within

Unity3D and consists of authentic artefacts, buildings, historical media content, and some surrounding cinematographic elements like wind and background music.

Figure 6 presents conceptual art of how the other episodes will look like in VR. In Figure 7, the concept of the connected storyline in VR is presented. The various locations of the six episodes are arranged along a path in the form of vignettes representing a virtual timeline. The user is able to move via teleportation from one episode to another depending on its interest.

3.1 The Proof-of-Concept

Due to the pandemic, the development of the complete experience was delayed. However, in May 2021, the first proof–of–concept of the VR experience "Ernst Grube – the legacy" was presented to the public in the NS Documentation Centre in Munich [22]. In this short 3:30 min. experience, Ernst Grube reflects about his stay in the Jewish children's home in Munich in 1940. This proof-of-concept

Figure 8: Pictures from the exhibition: (left) table with demo PC and cable hanger; (right) User in the VR Experience, on the wall there is a large display showing the current view of the user in the VR Experience.

gives an idea of this new concept and presents memories of a Holocaust survivor using volumetric video for the first time.

In mid 2022, it is planned to create the remaining five virtual scenes and include the volumetric video of Ernst Grube and the young student in there. The total duration of the VR experience will be approximately 50 min., whereas the user can switch between the different episodes and watch freely, depending on her/his desires and interests. In the context of cultural heritage, this VR experience will be the one with the longest duration using digital representations of humans based on volumetric video technology.

In Figure 8, some impressions from the exhibit in Munich are shown. In the delimited area, the user could move freely. The headset was connected to the render PC via a cable duct on the ceiling. At the same time, the current view of the users in the VR scene was shown on a large display (Figure 8, right photo below on the wall). This also gave other visitors the opportunity to get to know the VR experience.

3.2 First User Evaluation Results

With this proof-of-concept, it was possible to perform evaluations and to gather feedback. A questionnaire was laid out next to the VR experience to invite visitors to provide their feedback and perception. To the best of our knowl-edge, this is the first user evaluation of a VR experience using volumetric video technology in the context of preservation of memories.

A total number of 40 visitors took part in the evaluation. The gender distribution was 24:16 (female:male) with an equal age distribution ranging from young visitors to elderly. It is important to note that more than 70 % of the users did have no or just little previous experiences with Virtual Reality. This fact distributed equally among the different age groups. An important question is how the users exploits the possibility to navigate with 6 degrees of freedom in the virtual scene. Figure 9, left shows that a large proportion dared to move around in VR. More than a third of the users even approached the protagonists closely.

Furthermore, it was asked how the representation of the volumetric characters, the type of representation in relation to classical forms and the immersion was experienced. The diagram in Figure 9, right shows the results for the three questions on this topic. It is gratifying that the majority of the three questions were answered positively. The question about the special type of representation in relation to classic forms was rated even more positively (mean value = 3.68).

Finally, the duration of the VR experience, the emotional touch, the recommendation and the overall impression were asked (see Figure 10). The majority of visitors perceived the VR Experience as too short. Nevertheless, the respondents stated that they were touched by Ernst
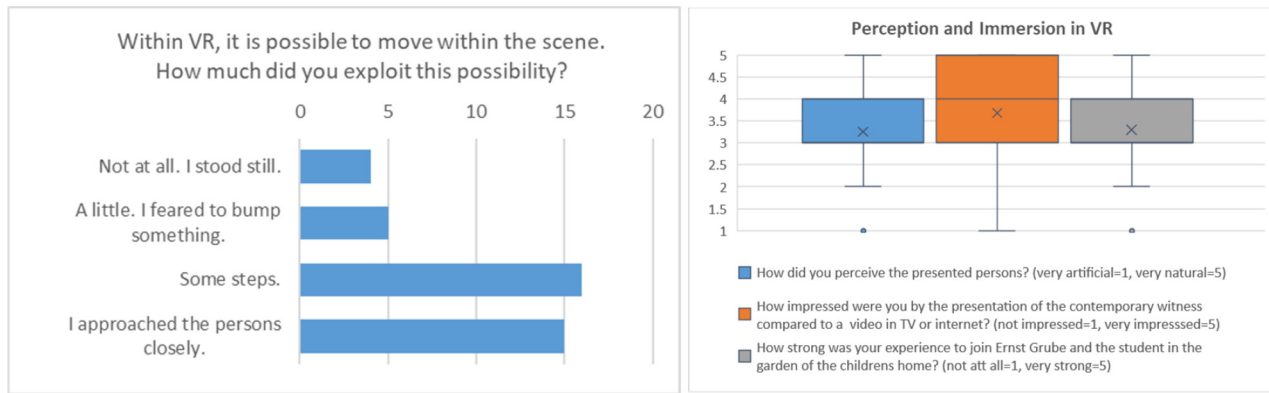
Figure 9: Evaluation results on movement in the virtual scene (left) and perception and immersion (right).
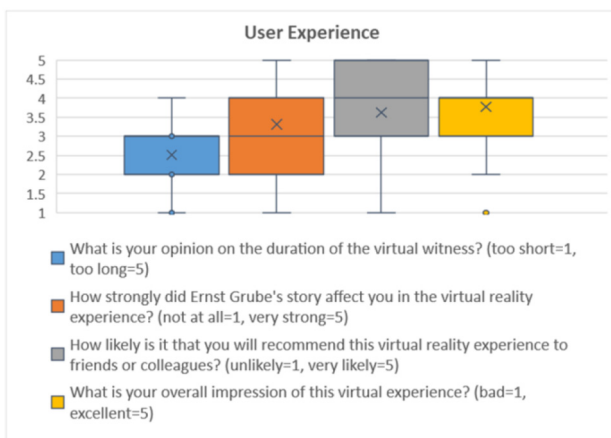

Figure 10: Evaluation results on user experience.

Grube's narration in the VR presentation (mean value = 3.32). This experience would be recommended by the majority to friends and colleagues (mean value = 3.63). Finally, the overall impression also tended to be rated positively (mean value = 3.78).

This user evaluation is the very first attempt to get user feedback for such kind of VR experience. We plan to perform more detailed user studies as soon as the final experience is available. We believe that user studies of such kind of immersive experiences are necessary due to the sensitivity of the topic.

4 Testimony of Eva Umlauf

Our second project is based on a collaboration with Prof. Anja Ballis from the faculty of languages and literature and Prof. Markus Gloe from Geschwister-Scholl-institute for political science, both at Munich University.

By beginning of 2021, we developed a concept for the story to be told by Eva Umlauf. She is considered as one of the youngest children that survived the concentration camp in Auschwitz. Based on her biographic memories [23] and intensive discussions, six episodes of her life were selected: (1) Novaky, the place of a Slovakian concentration camp and her place of birth, (2) Auschwitz, (3) Antisemitism after the Second World War, (4) Transmigration from Slovakia to Germany in 1966, (5) Physician in Munich, (6) Auschwitz speech in 2011.

After finalizing the contentual concept with Eva Umlauf and the team, a shooting was performed in May 2021 in the volumetric video studio at Volucap GmbH in Potsdam-Babelsberg (see some impressions in Figure 11). Each episode was planned to last between 8–10 minutes. Overall, almost 60 minutes of content were captured, leading to a total amount of 80 TeraBytes of raw video data. The concept and outlook of the virtual scene, where Eva Umlauf will be placed in, is currently under planning. The creation of volumetric video as well as the development of the complete VR experience is envisaged for 2022 to be realized thanks to a recently started national project.

Figure 11: Eva Umlauf in the volumetric studio.

Figure 12: Eva Umlauf: two different views, original (left), reconstructed meshes (middle), reconstructed and textured meshes (right).

However, some first results are already available. Figure 12 presents two different camera views of Eva Umlauf inside the capture stage. Figure 12 – left, shows the original camera view. Figure 12 – middle, shows the related geometry resulting from our volumetric video workflow. Finally, on the right, the textured mesh of the resulting volumetric asset is presented from the view of the original camera views on the left.

5 The User Perspective in VR

The immersive presentation of memories from survivors of the Holocaust requires a careful thinking of how this is realized in Virtual Reality, because of the importance and sensitivity of the topic. The user in VR has the opportunity to get close to the contemporary witness and observe its oral presentation of memories from arbitrary positions in the virtual space. Hence, it is required to develop a respect-

ful experience, which puts the user in a clear role to reach an emotional relationship with the contemporary witness. Currently, we investigate and follow two concepts about how the user may experience the memories of a contemporary witness.

5.1 Passive User Perspective

The first concept is a passive user approach, realized in the first production with Ernst Grube. Here, a young student serves as a counterpart to Ernst Grube by listening to his memories. As a user in VR, we join this conversation in a passive way, by just watching and listening the two protagonists. In Figure 13, left, the scenario is depicted, where the user in VR is marked in red. The arrows pointing from the nose of each actor assign the viewing direction. We choose a young student 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
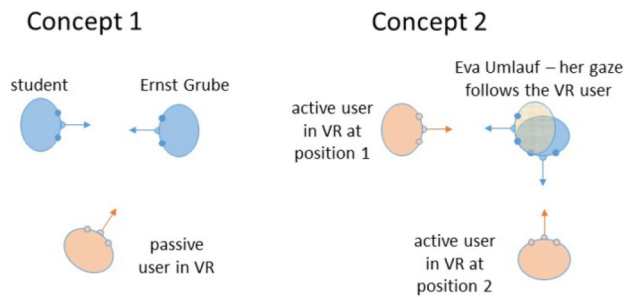
Concept 1

Concept 2

student   Ernst Grube

Eva Umlauf – her gaze
follows the VR user

active user
in VR at
position 1

passive
user in VR

active user
in VR at
position 2

Figure 13: Schematic description of user experience concepts.

this dark chapter of German history. The user follows the talk and can experience the emotions and gestures of Ernst Grube in a highly realistic 3-dimensional fashion.

5.2 Eye-Contact Preserving User Perspective

Nevertheless, what about if the contemporary witness is not talking to some other person in the scene, but to you as the user of the VR experience? This leads us to the second concept that is followed by the project "Testimony of Eva Umlauf". There is a fundamental problem, as Eva Umlauf is only looking in the direction, where she looked during acquisition. Hence, there is a need to modify the viewing direction of the volumetric character according to the position of the user in the VR experience to achieve eye contact. The two different viewing positions of the active user in VR and the envisaged head orientation of the volumetric character to provide eye contact are visualized in Figure 13, right. Due to recent advances in volumetric video, it is possible today to animate a volumetric character afterwards in the VR experience in real-time in the render engine [24] (see section 6 for more details).

With this technology, we are able to rotate the head, obviously within the given physiological limits, and make the volumetric character look at the user's eyes in the VR experience as illustrated in Figure 13, right.

5.3 Differences Between AR and VR

The two projects presented in the previous sections are dedicated to Virtual Reality applications. The reason is that the level of immersion should be increased through a three-dimensional scenery that is related to the content of the story told by the contemporary witness. This is further enriched by additional video material and additional historical information. Such a Virtual Reality application requires a costly VR head set and a related render engine.

However, this is not always available, especially if such kind of experiences shall be used in the context of history courses in secondary schools. The availability of Augmented Reality devices such as smartphones or tablets is more likely. Thus, the presentation of such use cases has several consequences for the application itself. First of all, the protagonist can be placed arbitrarily in the space and the user is able to follow his/her talk. By walking around the volumetric character, the three-dimensional nature of the protagonist can be experienced. The eye contact preserving scenario can also be easily transferred to AR, because the head orientation can be manipulated in the same way using the position of the AR device (either glasses, tablets or smartphones) as reference point. The major difference is that the user itself is not transferred to another virtual environment, because he/she stays in the real environment. It would be an interesting research question, what the user experience is for such use cases either presented in VR or in AR?

6 Interactive Animation of Volumetric Video

The key pre-requisite for the second concept is the possibility to generate animations and to apply modifications to the captured volumetric video stream. We address this by enriching the captured data with semantic information and animation properties by fitting a parametric kinematic human body model to it. The output of this step is a volumetric video stream with an attached parametric body model, which can be animated via an underlying skeleton model. During the fitting process, both shape and pose of the model are estimated. The shape is adapted from a template human body mesh to fit the character, and is kept constant for the complete sequence [25]. Care is taken to ensure temporal consistency of the poses over neighboring frames to avoid artifacts in the later animation stage. The result of the fitting process is a sequence of template meshes that are close to the meshes in the volumetric sequence, while lacking some of the finer grained details and textures.

For every vertex of each mesh in the volumetric sequence, we record the closest face of the template mesh and the closest point on this face in barycentric coordinates.

If a user joins the VR experience, head position and orientation is registered continuously by the 3D glasses. This information is used to estimate the correct perspective of the volumetric asset in real-time. A plugin in the ren-

Figure 14: Captured/original frame of the Volumetric Video data (1st image on the left) and animated (gaze corrected) frames (all remaining images on the right).

der engine exploits this positional information and feeds a render module that re-calculates a modified mesh for the current frame.

The previously fitted pose in the neck joints of the human body model are modified to turn the model's head towards the user. We define limits for every axis of the modified joints to ensure the resulting poses look natural, even if the user moves behind the character in the volumetric video stream.

The fitted template mesh is now modified to represent the new pose. As we previously recorded the closest points in the template mesh from the volumetric mesh, we can now modify and animate the original volumetric mesh sequence by moving it like the template mesh.

The modified mesh consists of a rotated head of the Volumetric Video character, i. e. Eva Umlauf, looking in the direction of the user. Thus, we keep all the geometric and texture details from the volumetric recording while modifying the pose to achieve a more immersive experience. In Figure 14, an example is given from a previous project, where the original gaze on the left was manipulated to the final orientation on the right. The head follows the new gaze direction.

7 Conclusion

In this paper, we presented two Virtual Reality projects, where volumetric video technology is used to recreate experiences with the contemporary witnesses. Both projects are one of the first in this domain, which benefit from high-quality dynamic 3D reconstruction of humans with volumetric video. Thanks to a first public presentation of the proof-of-concept of "Ernst Grube – the legacy", preliminary results were received from a small user study. The initial results are promising and encourage to further use volumetric video in such kind of genre.

We also presented two different concepts with regards to the user perspective in VR. The level of immersion is not yet investigated in the context of preservation of mem-

ories. Hence, the final VR experiences about Ernst Grube and Eva Umlauf will offer the capability to further research on this topic and create convincing, touching and immersive experiences in a highly sensitive historical context.

References

[1] Ballis, A., Gloe, M.: Interactive 3D Testimonies of Holocaust Survivors in German language. Springer Fachmedien Wiesbaden, Wiesbaden, 2020, pp. 343–368. doi:10.1007/978-3-658-24207-7_21.

[2] LMU: Learning with digital testimonies. https://www.en.lediz.uni-muenchen.de/projekt-lediz/index.html, 2021. Accessed: 21.06.2021.

[3] USC ICT: New dimensions in testimony. https://ict.usc.edu/prototypes/new-dimensions-in-testimony/, 2021. Accessed: 21.06.2021.

[4] Mori, M.: The Uncanny Valley (K. MacDorman & N. Kageki, Trans.). IEEE Robotics & Automation Magazine 19(2), 98–100 (2012). https://doi.org/10.1109/MRA.2012.2192811.

[5] W. Waizenegger, I. Feldmann, and O. Schreer: Real-time patch sweeping for high-quality depth estimation in 3D videoconferencing applications, In Proc. of Real-Time Image and Video Processing, San Francisco, California, United States, 2011.

[6] Volucap: http://www.volucap.de/.

[7] 4D VIEW SOLUTIONS SAS: Holosys. https://www.4dviews.com/holosys-volumetric-video-system, 2021.

[8] 8I: Hologram capture systems. https://www.8i.com/, 2021.

[9] Volograms: Reconstruction technology. https://volograms.com/technology, 2021.

[10] Vlasic D., Peers P., Baran I., Debevec P., Popovic J., Rusinkiewicz S., Matusik W.: Dynamic shape capture using multiview photometric stereo, In ACM SIGGRAPH Asia 2009 Papers, New York, NY, USA, 2009.

[11] Collet A., Chuang M., Sweeney P., Gillett D., Evseev D., Calabrese D., Hoppe H., Kirk A., Sullivan S.: High-quality streamable free-viewpoint video. ACM Trans. Graph. 34, 4 (July 2015).

[12] Guo K., Lincoln P., Davidson P., Busch J., Yu X., Whalen M., Harvey G., Orts-Escolano S., Pandey R., Dourgarian J., Tan D.,

Tkach A., Owdle A., Cooper E., Dou M., Fanello S., Fyffe G., Rhemann C., Taylor J., Debevec P., Izadi S.: The relightables: Volumetric performance capture of humans with realistic relighting. ACM Trans. Graph. 38, 6 (Nov. 2019).

[13] Worchel, M., Zepp, M., Hu, W., Schreer, O., Feldmann, I., Eisert, P.: Ernst Grube: A Contemporary Witness and His Memories Preserved with Volumetric Video, Eurographics Workshop on Graphics and Cultural Heritage (GCH 2020), Online Conference, 2020, ISBN: 978-3-03868-110-6, October 2020, doi:10.2312/gch.20201300.

[14] Horprasert T., Harwood D., Davis L. S.: A statistical approach for real-time robust background subtraction and shadow detection, In IEEE Int. Conf. on Computer Vision (ICCV 1999) Frame-Rate Workshop, 1999, 1999.

[15] Lin, S., Ryabtsev, A., Sengupta, S., Curless, B., Seitz, S., Kemelmacher-Shlizerman, I.: Real-Time High-Resolution Background Matting, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 8762–8771.

[16] W. Waizenegger, I. Feldmann, O. Schreer, P. Kauff, and P. Eisert: Real-time 3D body reconstruction for immersive TV, In 2016 IEEE International Conference on Image Processing (ICIP), 2016, pp. 360–364.

[17] Curless B., Levoy M.: A volumetric method for building complex models from range images, In Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, New York, NY, USA, 1996, SIGGRAPH '96, Association for Computing Machinery, pp. 303–312. doi:10.1145/237170.237269. 2, 3.

[18] Newcomber R. A., Izadi S., Hilliges O., Molyneaux D., Kim D., Davison A. J., Kohi P., Shotton J., Hodges S., Fitzgibbon A.: Kinectfusion: Real-time dense surface mapping and tracking, In 2011 10th IEEE International Symposium on Mixed and Augmented Reality, 2011, pp. 127–136. doi:10.1109/ISMAR.2011.6092378. 3.

[19] Kazhdan M., Hoppe H.: Screened poisson surface reconstruction. ACM Trans. Graph. 32, 3 (July 2013).

[20] R. Diaz, A. Shehu, I. Feldmann, O. Schreer, P. Eisert: Region Dependent Mesh Refinement for Volumetric Video Workflows, In Proc. Intern. Conference on 3D Immersion (IC3D), Brussels, Belgium, Dec. 2019.

[21] W. Morgenstern, A. Hilsmann, P. Eisert: Progressive Non-rigid Registration of Temporal Mesh Sequences, In Proc. European Conference on Visual Media Production (CVMP), London, UK, Dec. 2019, pp. 1–10.

[22] https://www.ns-dokuzentrum-muenchen.de/en/home/.

[23] Umlauf, E. & Oswalt, S.: Die Nummer auf deinem Unterarm ist blau wie deine Augen. Hoffmann und Campe, Hamburg, 2016, ISBN 978-3-455-50370-8.

[24] Hilsmann, A., Fechteler, P., Morgenstern, W., Paier, W., Feldmann, I., Schreer, O., Eisert, P.: Going beyond Free Viewpoint: Creating Animatable Volumetric Video of Human Performances, IET Computer Vision, Special Issue on Computer Vision for the Creative Industries, May 2020, doi:10.1049/iet-cvi.2019.0786.

[25] P. Fechteler, A. Hilsmann, P. Eisert: Markerless Multiview Motion Capture with 3D Shape Model Adaptation. Computer Graphics Forum 38(6), 91–109 (Sep. 2019).

Bionotes

Oliver Schreer
Fraunhofer Institute for Telecommunications
Heinrich-Hertz-Institute HHI, Vision & Imaging Technologies, Einsteinufer 37, 10587 Berlin,Germany
oliver.schreer@hhi.fraunhofer.de

Dr. Oliver Schreer is head of research group at "Vision & Imaging Technologies" department at Fraunhofer HHI and Associate Professor at TU Berlin. He published more than 120 papers. In May 2019, Oliver Schreer and his colleagues I. Feldmann and P. Kauff received the Joseph-von-Fraunhofer Prize entitled by "Realistic people in virtual worlds - A movie as a true experience".

Markus Worchel
Fraunhofer Institute for Telecommunications
Heinrich-Hertz-Institute HHI, Vision & Imaging Technologies, Einsteinufer 37, 10587 Berlin,Germany
markus.worchel@hhi.fraunhofer.de

Markus Worchel is a researcher in the Vision & Imaging Technologies department at Fraunhofer HHI. His research currently focuses on 3D reconstruction and material acquisition using differentiable rendering, implicit neural representations, and deep neural networks.

Rodrigo Diaz
Fraunhofer Institute for Telecommunications
Heinrich-Hertz-Institute HHI, Vision & Imaging Technologies, Einsteinufer 37, 10587 Berlin,Germany
RodrigoMauricio.DiazFernandez@hhi.fraunhofer.de

Rodrigo Diaz is a researcher at the Immersive Media & Communication group in the Vision & Imaging Technologies department at Fraunhofer HHI and a Ph.D. student at Queen Mary University in London. His current research interests include 3D reconstruction using deep learning methods, digital signal processing, virtual reality, and neural synthesis for audio.

Sylvain Renault
Fraunhofer Institute for
Telecommunications
Heinrich-Hertz-Institute HHI, Vision &
Imaging Technologies, Einsteinufer 37,
10587 Berlin,Germany
sylvain.renault@hhi.fraunhofer.de

Sylvain Renault is research associate at Immersive Media & Communications Group at Fraunhofer HHI. His main research field is the development of interactive XR applications. With focus on Unity3D, he has good experience with VR, HoloLens and AR mobile devices. He contributed to a multitude of national, European and industrial projects.

Marcus Zepp
Fraunhofer Institute for
Telecommunications
Heinrich-Hertz-Institute HHI, Vision &
Imaging Technologies, Einsteinufer 37,
10587 Berlin,Germany
marcus.zepp@hhi.fraunhofer.de

Marcus Zepp is a researcher in the Vision & Imaging Technologies department at Fraunhofer HHI. His research topics include point cloud fusion for 3D human body reconstruction, simultaneous localization and mapping, implicit representations and semantic segmentation with deep neural networks.

Wieland Morgenstern
Fraunhofer Institute for
Telecommunications
Heinrich-Hertz-Institute HHI, Vision &
Imaging Technologies, Einsteinufer 37,
10587 Berlin,Germany
wieland.morgenstern@hhi.fraunhofer.de

Wieland Morgenstern is a researcher in the Computer Vision & Graphics group in the Vision & Imaging Technologies department at Fraunhofer HHI. His research focuses on human body model registration and animatable volumetric video.

Anna Hilsmann
Fraunhofer Institute for
Telecommunications
Heinrich-Hertz-Institute HHI, Vision &
Imaging Technologies, Einsteinufer 37,
10587 Berlin,Germany
anna.hilsmann@hhi.fraunhofer.de

Anna Hilsmann is heading the Vision & Imaging Technologies Department as well as the Computer Vision & Graphics Group at Fraunhofer HHI. She received her Dipl.-Ing. degree in Electrical Engineering and Information Technology from RWTH Aachen in 2006 and her Dr.-Ing. degree in Computer Science from HU Berlin, in 2014. Her main research interests cover computer vision, computer graphics, deep learning and visual computing.

Ingo Feldmann
Fraunhofer Institute for
Telecommunications
Heinrich-Hertz-Institute HHI, Vision &
Imaging Technologies, Einsteinufer 37,
10587 Berlin,Germany
ingo.feldmann@hhi.fraunhofer.de

Ingo Feldmann is head of the Immersive Media & Communication group at Fraunhofer Heinrich Hertz Institute. His major research interests and working fields cover the topics VR/AR/XR technologies, Volumetric Video, 3D reconstruction, AI based 2D/3D image processing and real-time 3D video. He acts as reviewer for major Computer Vision conferences.

Peter Eisert
Fraunhofer Institute for
Telecommunications
Heinrich-Hertz-Institute HHI, Vision &
Imaging Technologies, Einsteinufer 37,
10587 Berlin,Germany
peter.eisert@hhi.fraunhofer.de

Peter Eisert is Professor for Visual Computing at Humboldt University Berlin and heading the Vision & Imaging Department at Fraunhofer HHI. He has published more than 200 papers and Associate Editor of JIVP and JVIS. His research interests include 3D image analysis and synthesis, face body processing, computer vision, computer graphics, as well as deep learning.