Volumetric Video - Acquisition, Compression, Interaction and Perception

Eduard Zell, Fabien Castan, Simone Gasparini, Anna Hilsmann,

Michael Kazhdan, Andrea Tagliasacchi, Dimitrios Zarpalas, Nikolaos Zioulis

Abstract

Volumetric video, free-viewpoint video or 4D reconstruction refer to the process of reconstructing 3D content over time using a multi-view setup. This method is constantly gaining popularity both in research and industry. In fact, volumetric video is more and more considered to acquire dynamic photorealistic content instead of relying on traditional 3D content creation pipelines. The aim of the tutorial is to provide an overview of the entire volumetric video pipeline. Furthermore, it presents existing projects that may serve as a starting point to this topic at the intersection of computer vision and graphics.

The first part of the tutorial will focus on the process of computing 3D models from captured videos. Topics will include content acquisition with affordable hardware, photogrammetry, and surface reconstruction from point clouds. A remarkable contribution of the presenters to the graphics community is that they will not only provide an overview of their topic but have in addition open sourced their implementations. Topics of the second part will focus on usage and distribution of volumetric video, including data compression, streaming or post-processing like pose-modification or seamless blending. The tutorial will conclude with an overview of perceptual studies focusing on quality assessment of 3D and 4D content.

1. Course Overview

Difficulty: Beginner to Intermediate

Format: Half-day

Organizer: Eduard Zell (University of Bonn)

The iconic holograms from Star Wars inspired researchers and engineers to develop systems that extend the 2D space of video. But we are only now reaching the state where volumetric video is increasingly adopted by media productions. The difficulty in capturing and distributing volumetric content is that a 3D scene must be reconstructed for every frame. In addition, file formats, streaming and editing of the recorded content must be re-designed to incorporate the additional dimension. Proposed technical solutions are often closely linked to applications. While for teleconferencing real-time framerates are essential [TDL*18], high emphasis will be put on quality for offline productions where volumetric video is considered as an alternative for 3D content creation [GLD*19].

3D reconstruction from a single camera view is highly ill-posed. Even though, single camera reconstruction methods are constantly improving [PTY*19, LXS*20], the gap between training data and captured content may lead to incorrect reconstruction results, which becomes especially visible if the actors wear extraordinary dresses, costumes or hair cuts. Therefore, a multi-view setup is still preferred when accuracy and high resolution matter. Existing multi-view setup systems vary largely and can consist between a few consumer RGB [PAM*18] or RGBD cameras [KND15, SKP*18, HLC*18] and go up to 100 high-resolution video cameras in combination with a light stage and infrared projectors [JSL*19, CCS*15, EFR*17, GLD*19]. For such high-end solutions the amount of captured data, which can easily reach hundreds of gigabytes for short sequences, sets high hardware requirements for storage, network bandwidth and processing power to convert the recorded images to 3D scenes.

Formats for volumetric video vary largely and ranges from point clouds [MBC17], 3D meshes [CCS*15, PKC*17] or a volumetric representation e.g. encoded by a signed distance function (SDF) [TDL*18]. For solid objects, like human bodies, 3D meshes with a consistent vertex set over a few frames is often preferred over point clouds due to higher compression ratios [PKC*17] and higher visual quality [ZOGS20]. At the same time a mesh representation is not always ideal, especially for thin objects or hair. Finally, we must also look at editing solutions [PKC*16,PKCH18,MRS*21] to develop the full potential of a new media format. A probably unique property of volumetric video is that it can be considered as both, a linear medium like cinema or theater or an interactive medium similar to games, where a seamless transition between short animation cycles simulates life-like interactions [PKC*16, HFM*20].

The previous, highly compact description of existing volumetric video pipelines illustrates well the great number of elements required to convert recorded content into a volumetric video. At the same time it shows well the interaction of algorithms from computer vision, computer graphics and geometry processing, includ-

DOI: 10.2312/egt.20211035

ing the side-effect that recent developments in one domain may take some time to be adopted by the other field.

This tutorial has been designed with the following two goals in mind: First, it should provide an overview of aspects required to capture and distribute volumetric content and be a starting point for related literature and open sourced projects. The second central aspect of this tutorial is to raise awareness for related open source projects to facilitate replicating volumetric video pipelines, get access to an implementation of topics described in papers or textbooks, and encourage to contribute to open source projects to increase the personal reputation within the field.

2. Schedule

Part I - Capturing Volumetric Video with Open-Source Tools

- Photogrammetry pipeline
  (Fabien Castan and Simone Gasparini – 25min)
- Low-cost volumetric video with consumer grade sensors
  (Dimitris Zarpalas and Nikolaos Zioulis - 25min)
- Poisson surface reconstruction
  (Misha Kazhdan - 25min)

Part II - Beyond Capturing

- 4D compression and streaming
  (Andrea Tagliasacchi - 20min)
- Interactive volumetric videos
  (Anna Hilsmann - 30min)
- Perceptual aspects on volumetric video quality
  (Eduard Zell - 20min)

3. Lecturer Biographies

Fabien Castan is an R&D engineer at Technicolor Production Services. He is specialized in computer vision for Visual Effects. Graduated from IMAC engineering school (Image Multimedia Audiovisual and Communication), he previously worked at Duran Duboi, Ubisoft Motion Pictures and Mikros Image. He has worked on several research projects (French ANR and European projects) in the field of photogrammetry.

Simone Gasparini is an assistant professor at Toulouse INP and a member of the REVA research team at IRIT in Toulouse, France. He holds an M.Sc. in computer engineering from the Politecnico di Milano and a Ph.D. in Information Technology from the same institution. His research domain is Computer Vision, with a particular focus on structure from motion, 3D reconstruction, augmented reality and camera models and geometry.

Dimitrios Zarpalas is a senior researcher (grade C) at the Information Technologies Institute (ITI) of the Centre for Research and Technology Hellas (CERTH). He holds the diploma of Electrical and Computer Engineer from Aristotle University of Thessaloniki, A.U.Th, an MSc in computer vision from The Pennsylvania State University, and a PhD in medical informatics (Health Science School, department of Medicine, A.U.Th). He joined ITI in 2008, as an Associate Researcher. His main research interests are on 3D/4D computer vision and machine learning, volumetric video, 4D reconstruction of moving humans, their compression and transmission in real-time; 3D motion capturing, analysis and evaluation; 3D medical image processing and shape analysis of anatomical structures.

Nikolaos Zioulis is an R&D engineer collaborating with the Information Technologies Institute (ITI) of the Centre for Research and Technology Hellas (CERTH). He holds the diploma of Electrical and Computer Engineer from the Aristotle University of Thessaloniki, (A.U.Th), and is currently pursuing his PhD jointly at CERTH/ITI and Universidad Politécnica de Madrid (UPM). His research interests lie at the intersection of computer vision, computer graphics and machine learning. He is also the technical lead of the low-cost Volumetric Capture platform developed by the Visual Computing Lab of CERTH/ITI.

Michael Kazhdan is a professor in the Computer Science Department at Johns Hopkins University. His research has focused on the challenge of surface reconstruction and considers the manner in which Stokes's Theorem can be used for distributed and out-of-core reconstruction of high-resolution models from data consisting of billions of points. He has also been working on problems in the domain of image-processing, developing efficient streaming algorithms for solving the large sparse linear systems associated with modeling terapixel images in the gradient-domain, and on problems in geometry processing, focusing on the evolution of signals on surfaces, as well as the evolution of the surfaces themselves.

Andrea Tagliasacchi is a staff research scientist at Google Brain and an adjunct faculty in the computer science department at the University of Toronto. His research focuses on 3D perception, which lies at the intersection of computer vision, computer graphics and machine learning. In 2018, he was invited to join Google Daydream as a visiting faculty and eventually joined Google full time in 2019. Before joining Google, he was an assistant professor at the University of Victoria (2015-2017), where he held the "Industrial Research Chair in 3D Sensing". His alma mater include EPFL (postdoc) SFU (PhD, NSERC Alexander Graham Bell fellow) and Politecnico di Milano (MSc, gold medalist).

Anna Hilsmann heads the research group on Computer Vision & Graphics at the Fraunhofer Heinrich-Hertz-Institute in Berlin, Germany, since 2015. She holds a diploma in Electrical Engineering from RWTH Aachen University and a PhD degree in Computer Science (with distinction) from Humboldt University of Berlin. Her research focuses on 2D/3D image and video analysis and synthesis covering the whole processing chain from capturing, image-and video analysis and understanding to modelling and rendering, and lies at the intersection of computer vision, computer graphics and machine learning.

Eduard Zell is heading the research group on 4D Crop Reconstruction at the University Bonn (Germany), since summer 2020. Previously, we worked on volumetric video as well as character creation, perception and animation in academia or industry. He is recipient of the Eurographics PhD Award and the best thesis of the faculty award (Bielefeld University). Previous positions and educational background include Trinity College Dublin (Ireland), KAIST (South Korea) and Bournemouth University (UK).

4. Extended Abstracts

4.1. Photogrammetry pipeline

The ability to quickly generate a 3D model of an object is a key problem in many applications, from the reverse engineering for 3D printing to the content creation for Mixed Reality applications and Visual Special Effects (VFX). In recent years, many solutions have been proposed to generate 3D models from images using the well known techniques of photogrammetry and stereo-vision.

In this tutorial we will present Meshroom, an open-source software for 3D reconstruction from images. We will introduce its graphical interface and show how you can use it. We will present an example of usage in production in the context of VFX. We will then detail the photogrammetry pipeline regarding both the sparse and the dense part of the reconstruction pipeline. In the former, we will address the steps, from the feature extraction and matching to the Structure-from Motion to estimate the camera poses and generate the sparse point cloud of the scene. For the dense reconstruction part, we will present the algorithms for estimating the depth maps, generating the mesh surface and finally texture it. We will also present some good practices and recommendations for the images acquisition as it is fundamental for the quality of the final model.

Meshroom: github.com/alicevision/meshroom

4.2. Low-cost volumetric video with consumer grade sensors

While high-end setups currently support different facets of volumetric capture technology applications like content creation and live telepresence, there is a need to transition towards lower cost, portable setups for digitizing human performances, which are also more suitable for experimentation and accessible research. To encourage progress towards this, a low-cost Volumetric Capture system [SKP*18] was developed and made openly available along with documentation covering both its software and hardware aspects.

In this tutorial, apart from presenting and documenting its operational details, we will offer insights for its various design choices, the challenges that were encountered, and the lessons learned. These span both hardware and software topics like the selection of sensors, the balancing of processing power and costs, the design of the system, and even tripod mounts to ensure the portability of the setup. The presented system has been remotely and on-site deployed in various places around the globe as one of its central design goals apart from affordability was its portability and usability. The tutorial will also describe how these two important components for the aforementioned aspects have been addressed, namely the spatial [SDT*20] (*i.e.* StructureNet) and temporal alignment of the sensors, in ways that are sensor-agnostic – allowing for hybrid-sensor deployments – and scalable, offering flexible capturing scenarios. However, as expected, these design choices come at a cost, and the limitations of this approach will also be presented and discussed. Finally, the tutorial will present the work that has been supported and enabled by this low-cost volumetric capture system, showcasing potential uses. More importantly, taking into account the ongoing data-driven revolution, such systems can be used for easily collecting volumetric datasets, with an example being HUMAN4D [CSB*20].

Volumetric Capture: vcl3d.github.io/VolumetricCapture/ StructureNet: vcl3d.github.io/StructureNet/

4.3. Poisson surface reconstruction

Reconstructing surfaces from scanned 3D points has been an ongoing research area for several decades, and has only become more important as commodity 3D scanners have become ubiquitous. Classically, approaches for surface reconstruction have approached the problem in one of three ways: (1) fitting a simplicial complex to the point cloud and labeling simplices as either *interior* or *exterior*; (2) evolving a base surface so that it fits the points; or (3) fitting an implicit function to the point cloud and extracting an appropriate level-set using algorithm like Marching Cubes [LC87].

In this tutorial we will review the Poisson Reconstruction method [KBH06, KH13], an implicit method that takes as its input an oriented point cloud and returns a surface that is robust to noise, non-uniform sampling, and registration error. The key idea is to show that an oriented input cloud can be treated as a sampling of the gradient of field of an implicit function (specifically, the indicator function) resulting in a method reducing the reconstruction problem to the solution of a Poisson equation. Using an adapted spatial data-structure to discretize the Poisson equation, the reconstruction problem can be solved in time that is linear in the size of the input. We will look at how the method can be refined by incorporating positional interpolation constraints. And, time-permitting, we will consider how Dirichlet boundary constraints can be incorporated [KCRH20] to place hard constraints on the location of the reconstructed surface.

Poisson Reconstruction: github.com/mkazhdan/PoissonRecon

4.4. 4D compression and streaming

We describe a realtime compression architecture for 4D performance capture that is orders of magnitude faster than previous state-of-the-art techniques, yet achieves comparable visual quality and bitrate. We note how much of the algorithmic complexity in traditional 4D compression arises from the necessity to encode geometry using an explicit model (i.e. a triangle mesh). In contrast, we propose an encoder that leverages an implicit representation (namely a Signed Distance Function) to represent the observed geometry, as well as its changes through time. We demonstrate how SDFs, when defined over a small local region (i.e. a block), admit a low-dimensional embedding due to the innate geometric redundancies in their representation. We then propose an optimization that takes a Truncated SDF (i.e. a TSDF), such as those found in most rigid/non-rigid reconstruction pipelines, and efficiently projects each TSDF block onto the SDF latent space. This results in a collection of low entropy tuples that can be effectively quantized and symbolically encoded. On the decoder side, to avoid the typical artifacts of block-based coding, we also propose a variational optimization that compensates for quantization residuals in order to penalize unsightly discontinuities in the decompressed signal. This optimization is expressed in the SDF latent embedding, and hence can also be performed efficiently. We demonstrate our compression/decompression architecture by realizing, to the best of our knowledge, the first system for streaming a real-time captured 4D performance on consumer-level networks [TDL*18].

In follow-up work [TSC*20] we compressed the TSDF further by relying on a block-based neural network architecture trained end-to-end. To prevent topological errors, we losslessly compress the signs of the TSDF, which also upper bounds the reconstruction error by the voxel size. To compress the corresponding texture, we designed a fast block-based UV parameterization, generating coherent texture maps that can be effectively compressed using existing video compression algorithms.

4.5. Towards Animations Volumetric Video

Photo-realistic modelling and rendering of humans is extremely important for virtual reality (VR) environments, as the human body and face are highly complex and exhibit large shape variability but also, especially, as humans are extremely sensitive to looking at humans. Further, in VR environments, interactivity plays an important role. While purely computer graphics modeling can achieve highly realistic human models, achieving real photo-realism with these models is computationally extremely expensive. Hence, more and more hybrid methods have been proposed in recent years. We will address the creation of high-quality animatable volumetric video content of human performances. Going beyond the application of free-viewpoint volumetric video, these methods allow re-animation and alteration of an actor's performance through (i) the enrichment of the captured data with semantics and animation properties and (ii) applying hybrid geometry- and video-based animation methods that allow a direct animation of the high-quality data itself instead of creating an animatable model that resembles the captured data [HFM*20]. Semantic enrichment and geometric animation ability can be achieved by establishing temporal consistency in the 3D data [MHE19], followed by an automatic rigging of each frame using a parametric shape-adaptive full human body model. We will especially cover geometry- and video-based animation approaches that combine the flexibility of classical CG animation with the realism of real captured data. For pose editing, we will address example-based animation methods that exploit the captured data as much as possible combined with kinematic animation of the captured frames to fit a desired pose.

These methods can be combined with neural animation approaches [PHE20, PHE21], to learn the appearance of regions that are challenging to synthesize, such as the teeth or the eyes, to fill in missing regions realistically in an autoencoder-based approach. We will cover the full pipeline from capturing and producing high-quality video content, over the enrichment with semantics and deformation properties for re-animation and processing of the data for the final hybrid animation. Our approach consist of a neural face model created from volumetric performances that provides a parametric representation of facial expression and speech. This neural model allows synthesizing detailed facial expressions and to interpolate faithfully between expressions. Based on this model, we present an example-based animation approach that can synthesize consistent face geometry and texture according to a low dimensional expression vector. The texture is represented as dynamic textures reconstructed from the volumetric data. Additionally, we trained an auto-regressive network to learn the dynamics of speech and disentangles style which enables us to animate visual speech directly from text/visemes and perform simple and fast facial animation based on a high-level description of the content and to capture and adjust the style of facial expressions by modifying the low dimensional style vector.

4.6. Perceptual aspects on volumetric video quality

Although, multi-view setups are generic to the captured content, they are dominantly used for human performances. Eye tracking experiments reveal that strong differences exist between the fixation time of different body parts suggesting that the head and upper torso should have substantially higher reconstruction quality than the remaining body parts [MLH*09]. General perceptual experiments on meshes and 4D sequences suggest that smooth deformations, both in spatial and temporal domain, are less noticeable than high-frequency noise. Interestingly not only the type of deformation, but also the surface of the model has a strong influence whether the artifacts will be noticeable or not. While inaccuracies are easily recognized on smooth surfaces the same inaccuracies may remain unnoticed on rough surfaces [VS11, CLL*13]. But not only the surface itself, but also the type of the representation seems to have a large impact on the perceived quality. Given the existing compression algorithms for point clouds and polygon meshes, mesh representations seem to achieve a better trade-off between perceived quality and compression rate [ZOGS20]. Finally, low level cues which have been studied extensively in the context of low polygon modelling and LOD creation will be considered as well [LRC*03].

References

[CCS*15] 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). doi:10.1145/2766945. 1

[CLL*13] CORSINI M., LARABI M. C., LAVOUÉ G., PETŘÍK O., VÁŠA L., WANG K.: Perceptual metrics for static and dynamic triangle meshes. *Computer Graphics Forum 32*, 1 (2013), 101–125. doi:https://doi.org/10.1111/cgf.12001. 4

[CSB*20] CHATZITOFIS A., SAROGLOU L., BOUTIS P., DRAKOULIS P., ZIOULIS N., SUBRAMANYAM S., KEVELHAM B., CHARBONNIER C., CESAR P., ZARPALAS D., KOLLIAS S., DARAS P.: Human4d: A human-centric multimodal dataset for motions and immersive media. *IEEE Access 8* (2020), 176241–176262. doi:10.1109/ACCESS.2020.3026276. 3

[EFR*17] EBNER T., FELDMANN I., RENAULT S., SCHREER O., EISERT P.: Multi-view reconstruction of dynamic real-world objects and their integration in augmented and virtual reality applications. *Journal of the Society for Information Display 25*, 3 (2017), 151–157. doi:https://doi.org/10.1002/jsid.538. 1

[GLD*19] GUO K., LINCOLN P., DAVIDSON P., BUSCH J., YU X., WHALEN M., HARVEY G., ORTS-ESCOLANO S., PANDEY R., DOURGARIAN J., TANG D., TKACH A., KOWDLE 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). doi:10.1145/3355089.3356571. 1

[HFM*20] 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 14, 6 (2020), 350 – 358. doi:10.1049/iet-cvi.2019.0786. 1, 4

[HLC*18] HUANG Z., LI T., CHEN W., ZHAO Y., XING J., LEGENDRE C., LUO L., MA C., LI H.: Deep volumetric video from very sparse multi-view performance capture. In *Proceedings of the European Conference on Computer Vision (ECCV)* (September 2018). 1

[JSL*19] JOO H., SIMON T., LI X., LIU H., TAN L., GUI L., BANERJEE S., GODISART T., NABBE B., MATTHEWS I., KANADE T., NOBUHARA S., SHEIKH Y.: Panoptic studio: A massively multiview system for social interaction capture. *IEEE Transactions on Pattern Analysis and Machine Intelligence 41*, 1 (2019), 190–204. doi:10.1109/TPAMI.2017.2782743. 1

[KBH06] KAZHDAN M., BOLITHO M., HOPPE H.: Poisson surface reconstruction. In *Proceedings of the Fourth Eurographics Symposium on Geometry Processing* (Goslar, DEU, 2006), SGP '06, Eurographics Association, p. 61–70. 3

[KCRH20] KAZHDAN M., CHUANG M., RUSINKIEWICZ S., HOPPE H.: Poisson surface reconstruction with envelope constraints. *Computer Graphics Forum 39*, 5 (2020), 173–182. doi:https://doi.org/10.1111/cgf.14077. 3

[KH13] KAZHDAN M., HOPPE H.: Screened poisson surface reconstruction. *ACM Trans. Graph. 32*, 3 (July 2013). doi:10.1145/2487228.2487237. 3

[KND15] KOWALSKI M., NARUNIEC J., DANILUK M.: Livescan3d: A fast and inexpensive 3d data acquisition system for multiple kinect v2 sensors. In *2015 international conference on 3D vision* (2015), IEEE, pp. 318–325. 1

[LC87] LORENSEN W. E., CLINE H. E.: Marching cubes: A high resolution 3d surface construction algorithm. *SIGGRAPH Comput. Graph. 21*, 4 (Aug. 1987), 163–169. URL: https://doi.org/10.1145/37402.37422, doi:10.1145/37402.37422. 3

[LRC*03] LUEBKE D., REDDY M., COHEN J. D., VARSHNEY A., WATSON B., HUEBNER R.: *Level of Detail for 3D Graphics*. Morgan Kaufmann, 2003. 4

[LXS*20] LI R., XIU Y., SAITO S., HUANG Z., OLSZEWSKI K., LI H.: Monocular real-time volumetric performance capture. In *European Conference on Computer Vision* (2020), Springer, pp. 49–67. 1

[MBC17] MEKURIA R., BLOM K., CESAR P.: Design, implementation, and evaluation of a point cloud codec for tele-immersive video. *IEEE Transactions on Circuits and Systems for Video Technology 27*, 4 (2017), 828–842. doi:10.1109/TCSVT.2016.2543039. 1

[MHE19] MORGENSTERN W., HILSMANN A., EISERT P.: Progressive non-rigid registration of temporal mesh sequences. In *European Conference on Visual Media Production* (New York, NY, USA, 2019), CVMP '19, Association for Computing Machinery. doi:10.1145/3359998.3369411. 4

[MLH*09] MCDONNELL R., LARKIN M., HERNÁNDEZ B., RUDOMIN I., O'SULLIVAN C.: Eye-catching crowds: Saliency based selective variation. In *ACM SIGGRAPH 2009 Papers* (New York, NY, USA, 2009), SIGGRAPH '09, Association for Computing Machinery. doi:10.1145/1576246.1531361. 4

[MRS*21] MOYNIHAN M., RUANO S., SMOLIC A., ET AL.: Autonomous tracking for volumetric video sequences. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (2021), pp. 1660–1669. 1

[PAM*18] PAGÉS R., AMPLIANITIS K., MONAGHAN D., ONDREJ J., SMOLIC A.: Affordable content creation for free-viewpoint video and vr/ar applications. *Journal of Visual Communication and Image Representation Volume 53* (2018), 192–201. doi:10.1016/j.jvcir.2018.03.012. 1

[PHE20] PAIER W., HILSMANN A., EISERT P.: Neural face models for example-based visual speech synthesis. In *European Conference on Visual Media Production* (New York, NY, USA, 2020), CVMP '20, Association for Computing Machinery. doi:10.1145/3429341.3429356. 4

[PHE21] PAIER W., HILSMANN A., EISERT P.: Example-based facial animation of virtual reality avatars using auto-regressive neural networks. *IEEE Computer Graphics and Applications* (2021), 1–1. doi:10.1109/MCG.2021.3068035. 4

[PKC*16] PRADA F., KAZHDAN M., CHUANG M., COLLET A., HOPPE H.: Motion graphs for unstructured textured meshes. *ACM Trans. Graph. 35*, 4 (July 2016). doi:10.1145/2897824.2925967. 1

[PKC*17] PRADA F., KAZHDAN M., CHUANG M., COLLET A., HOPPE H.: Spatiotemporal atlas parameterization for evolving meshes. *ACM Trans. Graph. 36*, 4 (July 2017). doi:10.1145/3072959.3073679. 1

[PKCH18] PRADA F., KAZHDAN M., CHUANG M., HOPPE H.: Gradient-domain processing within a texture atlas. *ACM Trans. Graph. 37*, 4 (July 2018). doi:10.1145/3197517.3201317. 1

[PTY*19] PANDEY R., TKACH A., YANG S., PIDLYPENSKYI P., TAYLOR J., MARTIN-BRUALLA R., TAGLIASACCHI A., PAPANDREOU G., DAVIDSON P., KESKIN C., IZADI S., FANELLO S.: Volumetric capture of humans with a single rgbd camera via semi-parametric learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2019). 1

[SDT*20] STERZENTSENKO V., DOUMANOGLOU A., THERMOS S., ZIOULIS N., ZARPALAS D., DARAS P.: Deep soft procrustes for markerless volumetric sensor alignment. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2020), pp. 818–827. doi:10.1109/VR46266.2020.00106. 3

[SKP*18] STERZENTSENKO V., KARAKOTTAS A., PAPACHRISTOU A., ZIOULIS N., DOUMANOGLOU A., ZARPALAS D., DARAS P.: A low-cost, flexible and portable volumetric capturing system. In *IEEE International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)* (2018), pp. 200–207. 1, 3

[TDL*18] TANG D., DOU M., LINCOLN P., DAVIDSON P., GUO K., TAYLOR J., FANELLO S., KESKIN C., KOWDLE A., BOUAZIZ S., IZADI S., TAGLIASACCHI A.: Real-time compression and streaming of 4d performances. *ACM Trans. Graph. 37*, 6 (Dec. 2018). doi:10.1145/3272127.3275096. 1, 4

[TSC*20] TANG D., SINGH S., CHOU P. A., HANE C., DOU M., FANELLO S., TAYLOR J., DAVIDSON P., GULERYUZ O. G., ZHANG Y., IZADI S., TAGLIASACCHI A., BOUAZIZ S., KESKIN C.: Deep implicit volume compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2020). 4

[VS11] VASA L., SKALA V.: A perception correlated comparison method for dynamic meshes. *IEEE Transactions on Visualization and Computer Graphics 17*, 2 (2011), 220–230. doi:10.1109/TVCG.2010.38. 4

[ZOGS20] ZERMAN E., OZCINAR C., GAO P., SMOLIC A.: Textured mesh vs coloured point cloud: A subjective study for volumetric video compression. In *Twelfth International Conference on Quality of Multimedia Experience (QoMEX)* (Athlone, Ireland, 2020), IEEE. 1, 4