Bringing stage events with real actors to your home in VR

A low complexity solution for immersive media streaming

Dr.-Ing. Joachim Keinert
Fraunhofer Institute for Integrated Circuits IIS, joachim.keinert@iis.fraunhofer.de

Tobias Jaschke
Fraunhofer Institute for Integrated Circuits IIS, tobias.jaschke@iis.fraunhofer.de

Hendrik Hawer
Mobile-Etage GmbH, hawer@mobile-etage.de

Andree Josef
Mobile-Etage GmbH, josef@mobile-etage.de

Dr.-Ing. Frank Dittrich
Chemnitz University of Technology, frank.dittrich@mb.tu-chemnitz.de

Prof. Dr.-Ing. Siegfried Fößel
Fraunhofer Institute for Integrated Circuits IIS, siegfried.foessel@iis.fraunhofer.de

Abstract

In this paper we demonstrate a novel method how to integrate stereo content of a live act into a virtual reality (VR) environment. This opens a new level of immersion, when watching the recording of a stage event using a head mounted display. In contrast to existing solutions, we offer proper depth perception, allow for a low complexity recording and distribution and still avoid any motion sickness caused by objects being close to the observer. To this end, we separate the digitization of the event location from the recording of the actor's performance. As a main contribution of the paper, we show how to achieve high visual quality when compositing the stereo recording showing the actors with the reconstructed 3D-model of the event location. To do so, we have created a tool that allows to simulate the 3D perception of the stereo video depending on the selected stereo parameters such as inter-camera distance and focal length. By these means, we can ensure a smooth transition between the stereo video and the 3D model of the event location, leading to superior visual quality. Numerous lessons learned conclude our paper showing how we overcame practical problems encountered when recording a real theater play.

Keywords

Stereo, Parallax, Actor, Live Act, Stage, Event Location, 3D-Reconstruction, Blender, Unity, Virtual Reality, Head Mounded Display, Broadcast, 360° Video, Motion Capture, Photogrammetry, LIDAR, Baseline, Focal Length, Synchronization

1 INTRODUCTION

Broadcasting stage events such as theater, music or circus performances enlarges the potential audience to those not being able to attend the live event. Moreover, it allows to reexperience the happening once being home again. Unfortunately, traditional media such as television or cinema do not provide the feeling of participation. Consequently, the broadcasted experience is less thrilling than the true event.

Head mounted displays for virtual reality promise to significantly reduce this gap between the broadcasted and the live event. By surrounding the spectator with a complete and realistic environment, the feeling of presence is much more convincing than with a 2D screen. While 360° video is often used for such purposes, it lacks parallax. It is defined as the displacement or difference in the apparent position of an object viewed along two different lines of sight (1). In other words, when changing the observation point, the visualized objects shift relatively to each other or get even occluded (see Figure 1). When this is not properly considered, close objects like surrounding chairs in an event location quickly cause motion sickness. Alternatively, the event can be completely modelled in a game engine using 3D computer graphics. Then the parallax is perfectly reproduced, but it is hard to integrate live actors into the experience. Often, either green-screen or even volumetric capture is required, or the actors are completely replaced by avatars, reducing the naturalness.

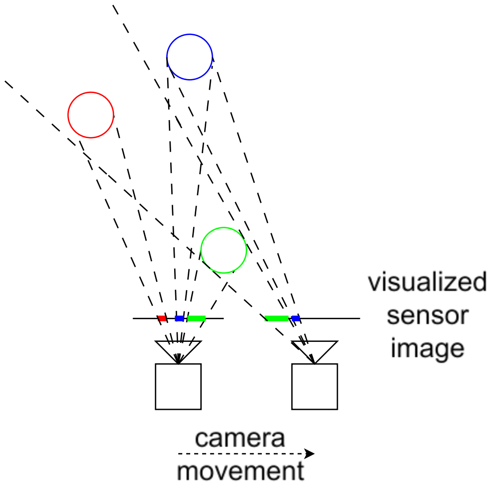


Figure 1: Impact of parallax. Objects seem to shift relatively to each other on the sensor image (green and blue circle) or even get occluded (red object), when the camera moves.

To overcome these drawbacks, the paper presents a novel workflow to integrate live actors into a 3D environment for fully immersive experience of stage events. To capture the atmosphere of the live event, the recording takes place directly at the event location, even with audience present. At the same time, complexity for recording, distribution and rendering is made as low as possible to allow for an economic application.

2 STATE OF THE ART

2.1 Cultural events in virtual reality

Cultural events can be represented in virtual reality using computer generated content only (2,3). This allows to generate cultural events using computer graphics software without involvement of any human actors on stage. Alternatively, performance of human actors can be recorded by motion capture systems (4,5), which rely on markers or camera-based computer vision software. By these means, the movements of the actors can be transferred to virtual avatars. While such technology has been proven extremely powerful in the visual effects industry, it is very tedious to digitize a complete stage event like concerts or theater plays with such a technology. This is, because not only the actors, but also involved objects such as instruments need to be faithfully digitized.

2.2 Photorealistic content capture for virtual reality

360° video capture is well-known to record (6,7), transmit (8,9) and edit (10) photorealistic content for head mounted displays (HMDs) (11,12), available both on professional (6,13) and consumer grade (7,14). However, 360° videos lack stereoscopic depth cues and only provides a correct perspective for a single observer position with an arbitrary head rotation. This quickly leads to motion sickness in case of close objects such as neighboring chairs. Omnidirectional stereo (15–18) improves this situation by providing a limited parallax (3DoF+). However, necessary view rendering (19) introduces the risk of artefacts, and the capture equipment is very expensive. Traditional stereo recording is well known for movie production (20–26), but combination with VR content raises new challenges that are not explored so far. Volumetric video (5,27–35) can provide all six degrees of freedom for movement (6DoF). However, it requires complex studio installations making it inappropriate for recording in event locations such as theaters or concert halls. Portable multi-camera arrays (36–43), also called light-field arrays (44), are more appropriate given the limited spectator movement for many cultural events such as theaters. However, view rendering (45–48), (27,39,49–56), (57–59) is prone to artefacts and computationally expensive.

3 SYSTEM OVERVIEW

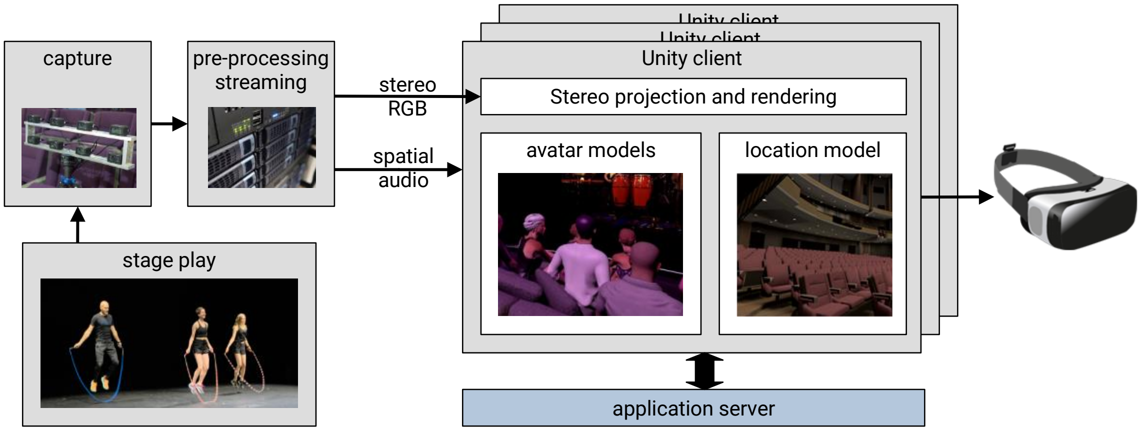


Figure 2: System diagram of the proposed solution.

Figure 2 illustrates our overall system. It visualizes a stage play in a 3D environment using an application server and multiple Unity clients. Each Unity client is serving one remote spectator wearing a VR headset. The actors on the stage are displayed in form of a stereo video for highest realism. Avatars represent the remote participants watching the live event. Additional computer-controlled avatars can visualize the live audience.

The following sections will detail the components related to the integration of the actors' performance into the VR experience, including recording, data processing, alignment of the stereo video with the 3D-model of the event location and visualization in Unity.

4 RECORDING OF STAGE EVENTS

4.1 Low Complexity Recording Solution

To reduce the complexity for recording, we decided to separate the recording of the actors' live act from the reproduction of the event location. For the latter, we build a 3D-model using photogrammetry and/or LIDAR scans (60,61) (see Figure 3). By these means, parallax of close objects can be perfectly reproduced. Imperfections not being properly reconstructed are manually cleaned by a 3D-artist (see Figure 4). Finally, geometry and texture simplification ensure sufficient rendering performance of the final Unity application.

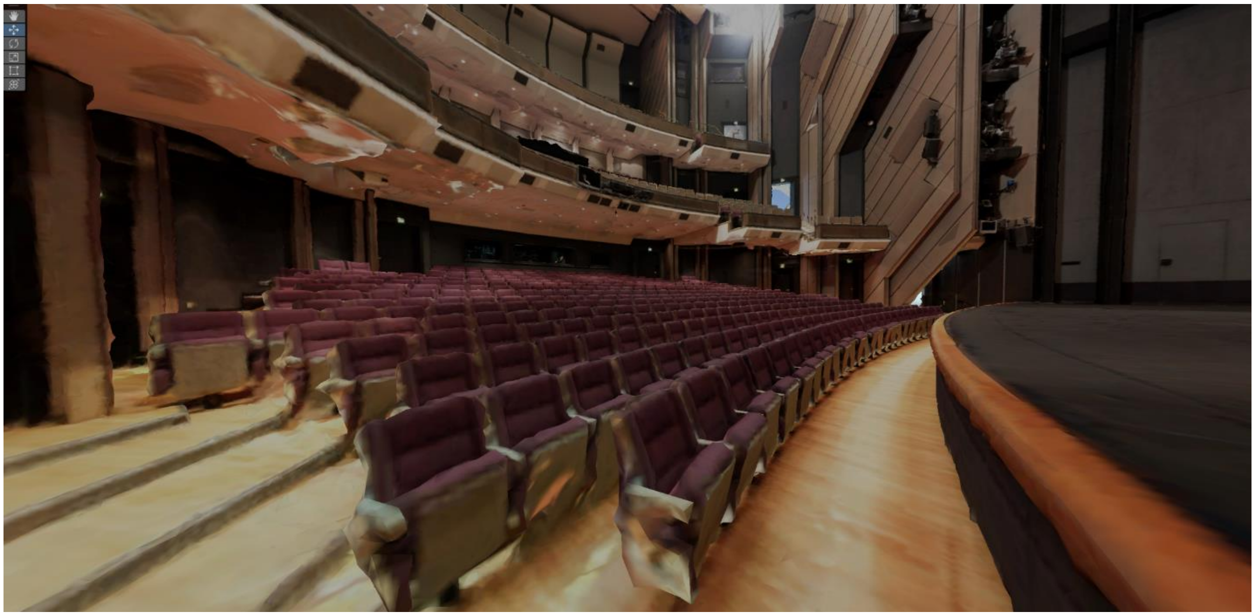


Figure 3: 3D model of the event location reconstructed from photogrammetry.

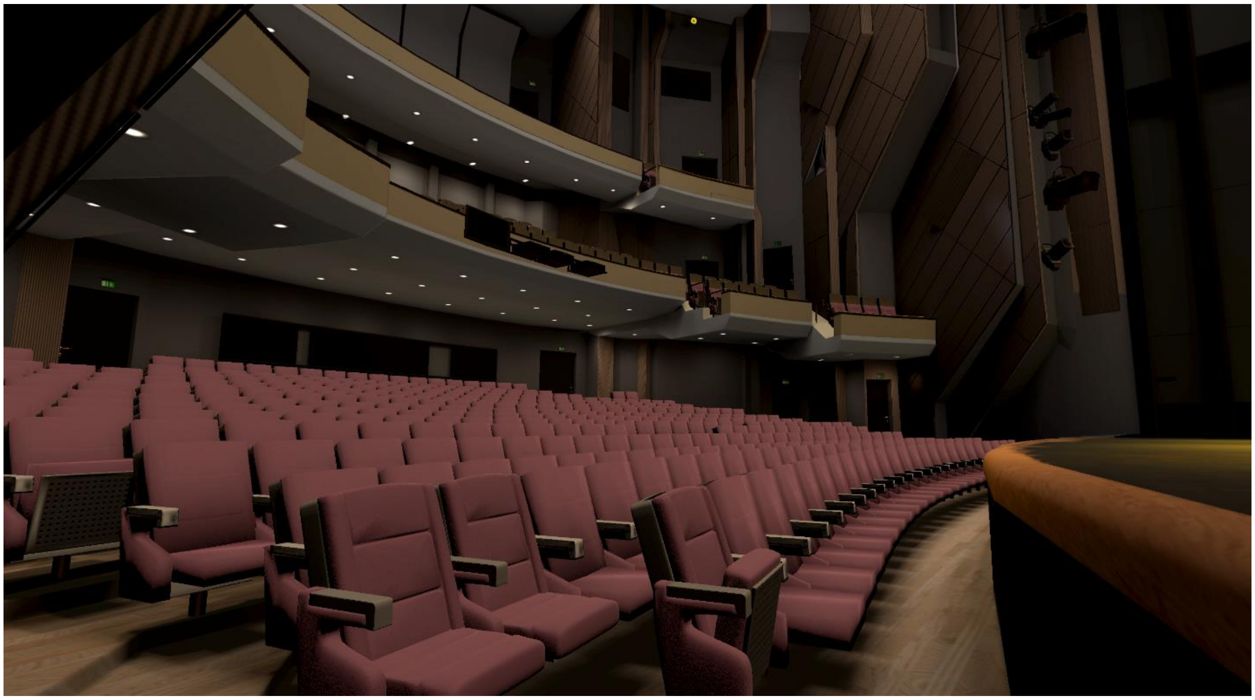


Figure 4: Cleaned and simplified 3D model of the event location shown in Figure 3.

For recording of the actors' live act, we first experimented with a light-field video capture (41,62). In theory it should give the best user experience, as we can synthesize the image of the live act corresponding exactly to the head position of the spectator. In practice, we found however that rendering artefacts negatively impact the experience. Moreover, compute times and number of required cameras were still too huge for a low complexity solution.

Consequently, we opted for a front-facing stereo camera to record the live-act. By these means, we can still visualize the actors' performance with the correct depth impression. Please note that stereo cameras only record a single perspective of the actors' performance. Consequently, there can be a mismatch between the perspective of the virtual observer and the one of the stereo cameras (see Figure 5). Fortunately, we found out that this effect is negligible when the distance between actors and spectators is sufficiently large, and when the stereo visualization is carefully designed in the VR application. More details will be described later.

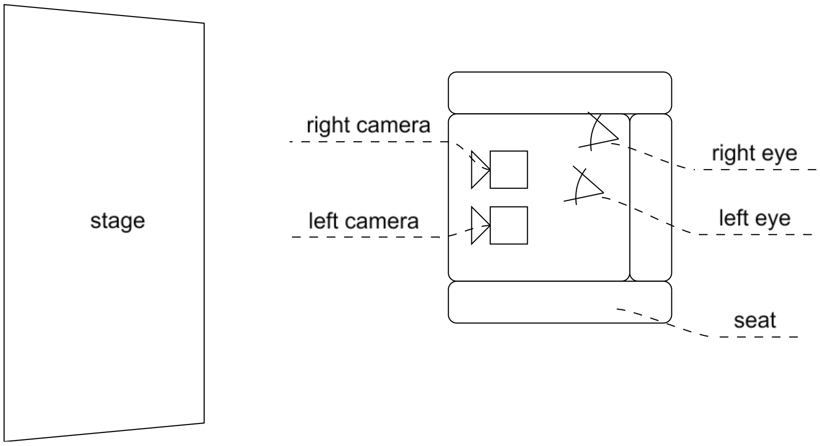


Figure 5: Mismatch of the perspective. Only when the eyes of the virtual observer match the location of the stereo camera, the visual impression is fully correct. Otherwise, visualizing the stereo images causes a parallax error.

4.2 Practical Problems and their Solutions

As our goal is to record the actual live event with possible audience on premise, the illumination is typically set to serve the latter. This results in many high-dynamic range and low-light conditions that are quite difficult to handle in practice with existing cameras. We started our experiments using Sony RX0-MK2 cameras (63), as they are small in size and can be synchronized. However, requiring ISO values between 320 and 1000 led to heavy image noise, which deteriorated the image quality to an inacceptable level. Cameras such as Sony-FX9 can be synchronized as well and deliver better quality. But in our experiments, we found that their body is too huge for the required stereo baseline (inter-camera distance). Using mirror rigs was not an option neither due to reducing the light intensity per camera by a factor of two (64). In our latest experiments we used Sony-FX3 cameras, as they combine acceptable quality with reasonable body sizes. We recorded with 25 frames per second to reduce light-flickering and have better sensitivity than for 50fps. We could compensate the missing physical synchronization of the FX3 cameras by up-sampling the framerate using an AI tool (65), followed by temporally aligning the left and right video stream as well as possible.

Figure 6: Stereo camera (left) placed in Theater of Chemnitz, Germany (right) for recording the actors.

While this worked remarkably well, it has to be noted that for future productions a solution with physical camera synchronization, small camera baseline and low image noise is highly desirable. A planed experiment with the Z-CAM E2 failed due to limited availability of a stereo pair. Luckily, with the advent of the Apple Vision Pro VR headset, more camera manufacturers have started to design corresponding setups (66).

5 INTEGRATION OF STEREO CONTENT INTO VR APPLICATIONS

5.1 Principle

To integrate the stereo videos in the overall VR experience, we project them on two virtual projection screens (see Figure 7). The left projection screen is visible in the left eye only, the right projection screen in the right eye only. While this sounds simple, we found it very difficult to manually place the screens to reach a good viewing experience. This is because wrong screen placement may cause the perceived depth from the stereo video to conflict with the depth from the 3D model of the event location. Then it may for instance happen that the actor visible in front of a wall may appear to be behind the wall, because the perceived depth is too large. This easily leads to VR sickness or eyestrain. Please note that this is a new challenge compared to traditional stereo movie production. In the latter, the stereo playback must only look pleasant, but does not need to fit with any virtual environment.

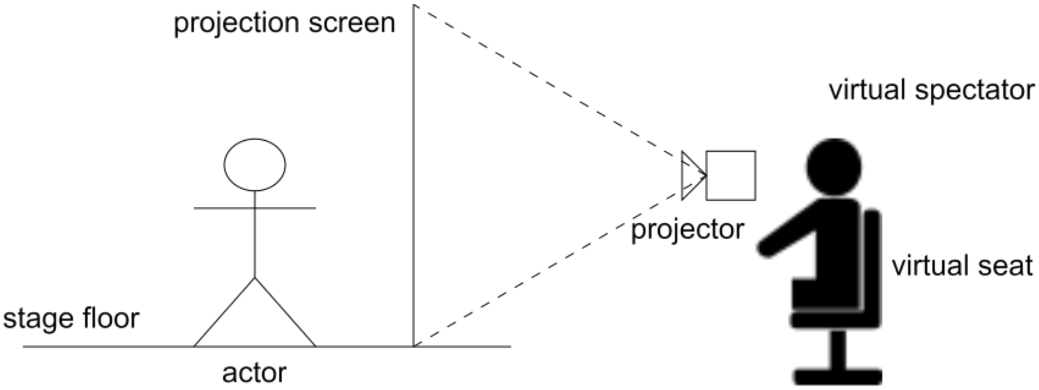


Figure 7: Integration of videos by projection on a virtual screen.

In theory the described problem can be easily avoided. Suppose to have a perfect 3D model of the event location. Then we simply need to setup two projectors at the same location than the stereo cameras during capture. Each of the projectors projects the captured video onto the virtual projection screens (see Figure 7). The intrinsic and extrinsic parameters of the recording stereo cameras and the projectors need to match perfectly. Intrinsic camera parameters include for instance the focal length and the position of the sensor center relative to the optical axis of the camera. Extrinsic camera parameters encompass locations and rotations of both cameras. Finally, the eyes of the human should be close to the projector locations.

In practice, none of these constraints can be easily met. First, we typically do not know intrinsic and extrinsic camera parameters with sufficient precision relative to the coordinate system of the 3D model of the event location. And while there exist algorithms to perform camera calibration (67), they only calibrate towards a relative coordinate system, and not towards the absolute one of the event location. Even worse, some camera parameters such as the position of the sensor center relative to the optical axis are difficult to determine reliably, although being crucial for the perceived depth impression. Moreover, practical constraints like camera body sizes prevent the two stereo cameras from having exact inter-eye distance. In addition, the location of the stereo cameras may deviate from the eye position of the virtual observer. And finally, for artistic reasons it may be desired to increase the perceived depth impression compared to the perfectly realistic reproduction.

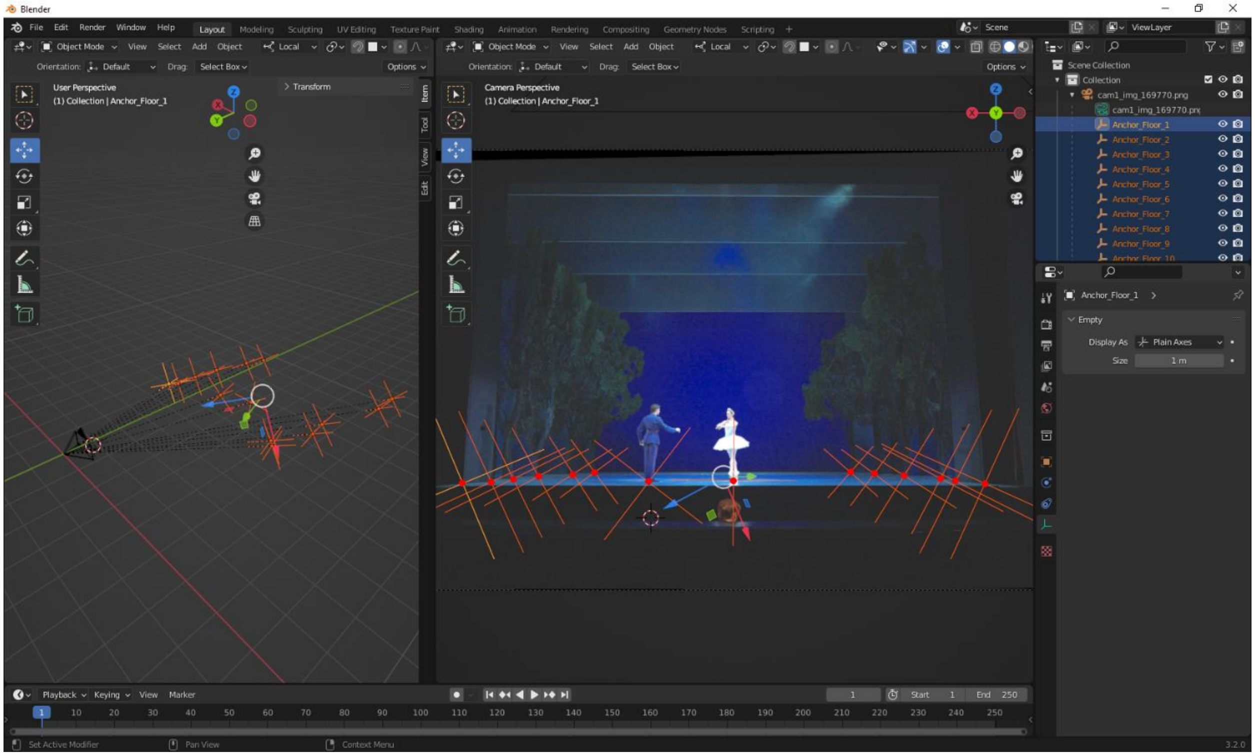
All these reasons cause that a cleverer solution is needed to integrate the stereo videos into the VR environment.

5.2 Solution

To do so, we have created a rapid prototype solution using Matlab and Blender. We first rectify the input videos using image correspondences (54,68). By these means, we can correct the images such that they correspond to two stereo cameras with perfectly parallel optical axes. This is important, because even with careful mechanical design of the stereo rig, imprecisions during manufacturing and assembly cause that such an alignment is hard to achieve with a precision of a single pixel.

While we tried to use checker boards to achieve a metric calibration, capturing them in practice showed to be difficult. First, most of them are too small to be useful when placed on the stage due to the large distance to the cameras. Second, they often interfere with the course of the events in the theater. Please note that despite the rectification, it stays important to align the stereo rig in such a way that the line interconnecting the nodal points of the lenses is approximately orthogonal to the direction of gravity (or more precisely, parallel to the line interconnecting the eyes of the observer). Otherwise, the resulting mismatch between the stereo video and the 3D model of the theater can only be corrected by rotating the theater model. This quickly looks odd. We used a spirit level to avoid this problem.

Then, we import the stereo video into Blender. To do so, we instantiate two cameras representing the stereo rig. Next, we assign each video to the corresponding camera. With the help of a custom plugin, we can then mark reference points in both the left and right image (see Figure 8). Then, each reference point is projected into the 3D space. By these means, we simulate where an observer located at the position of the stereo camera would see the reference point in 3D space. Typically, we define multiple reference points on the floor and the walls. We also define some reference points for actors. By these means, we can later control, that their perceived height is close to reality. Please note that this process may be speed-up by automatic feature matching (68) or even stereo depth estimation (69–71). However, to have better control over the matching quality, we stayed with manual selection for our rapid prototype.



Display As: Plain Axes

Size: 1 m

Figure 8: Definition of the reference points in Blender. On the right side, we look through one of the cameras. Red points identify reference points. On the left side we see the projected points in 3D space.

In our scenes, we found the floor to be most relevant for matching the stereo video with the 3D model of the theater. We want the floor visible in the stereo video to transition smoothly into the floor of the 3D theater model. Consequently, we fitted a corresponding plane to the floor of the stereo video using its 3D reference points. Other crucial scene elements such as pillars can be handled in a similar fashion.

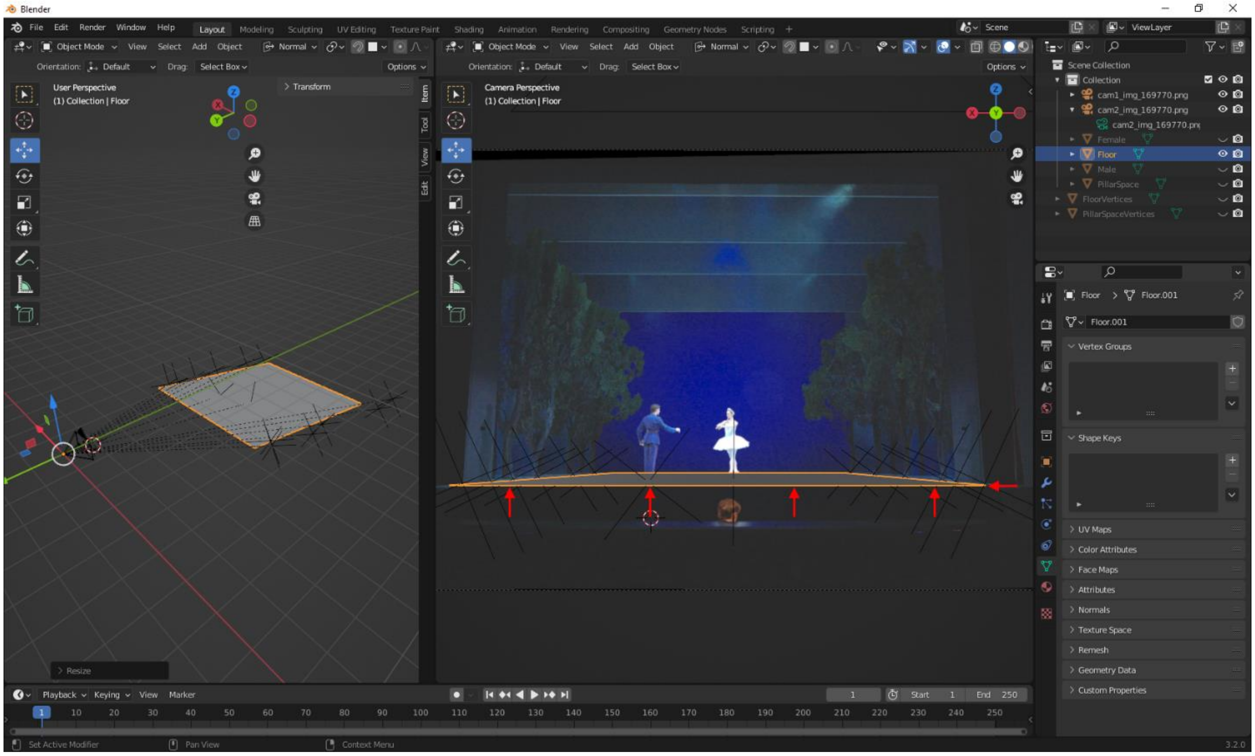


Figure 9: Fitting a plane to the floor of the stereo video.

Once we have covered the essential elements of the stage in the stereo video, a further custom Blender plugin allows us to simulate how changing the stereo parameters will impact the perceived impression by the virtual observers. In more detail, our Blender workflow allows to simulate the change of the stereo projector location and orientation, the stereo baseline, the focal length, the position of the sensor center relative to the optical axis, the sensor offset (difference of the sensor centers' positions between the two cameras in the direction of the baseline), the impact of a mismatch between the observer's eyes and the projectors, and the distance between the projection screen and the projectors (see Figure 10).

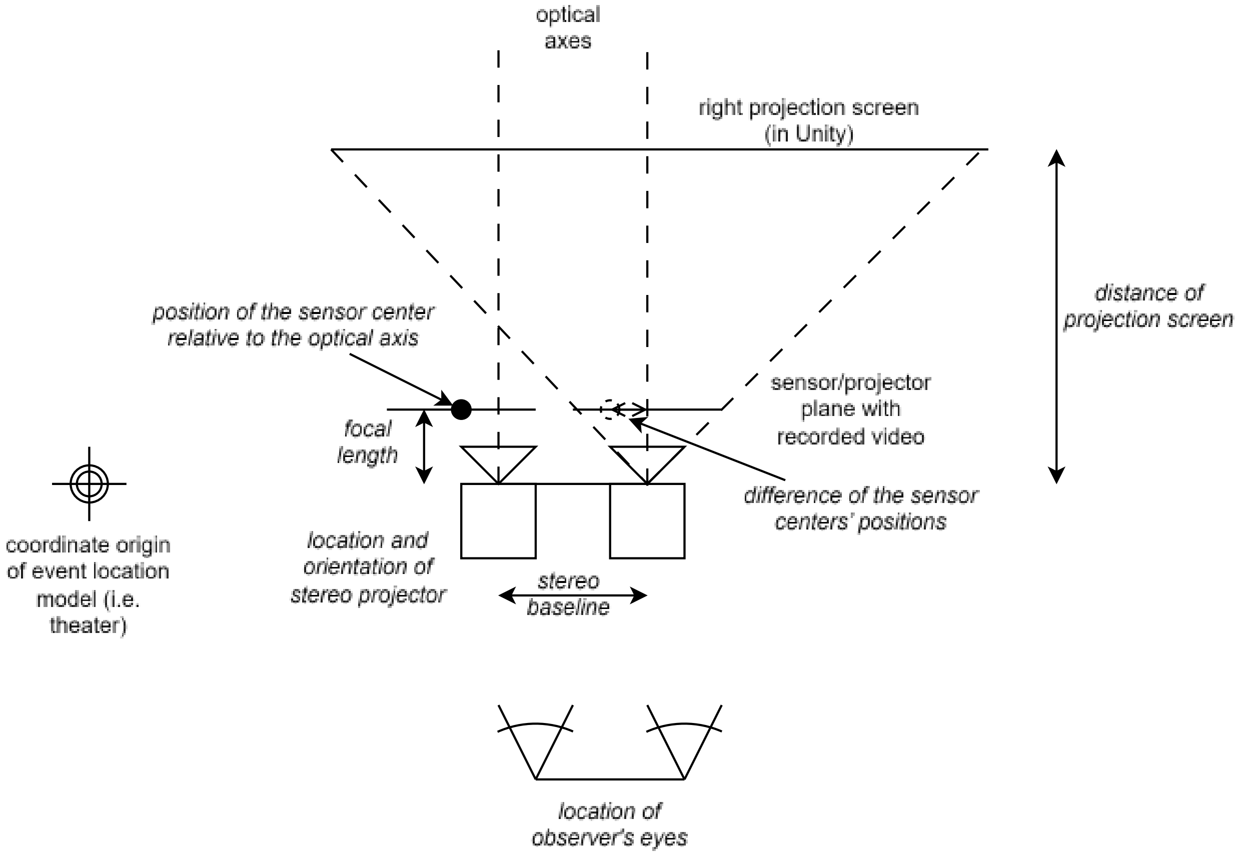
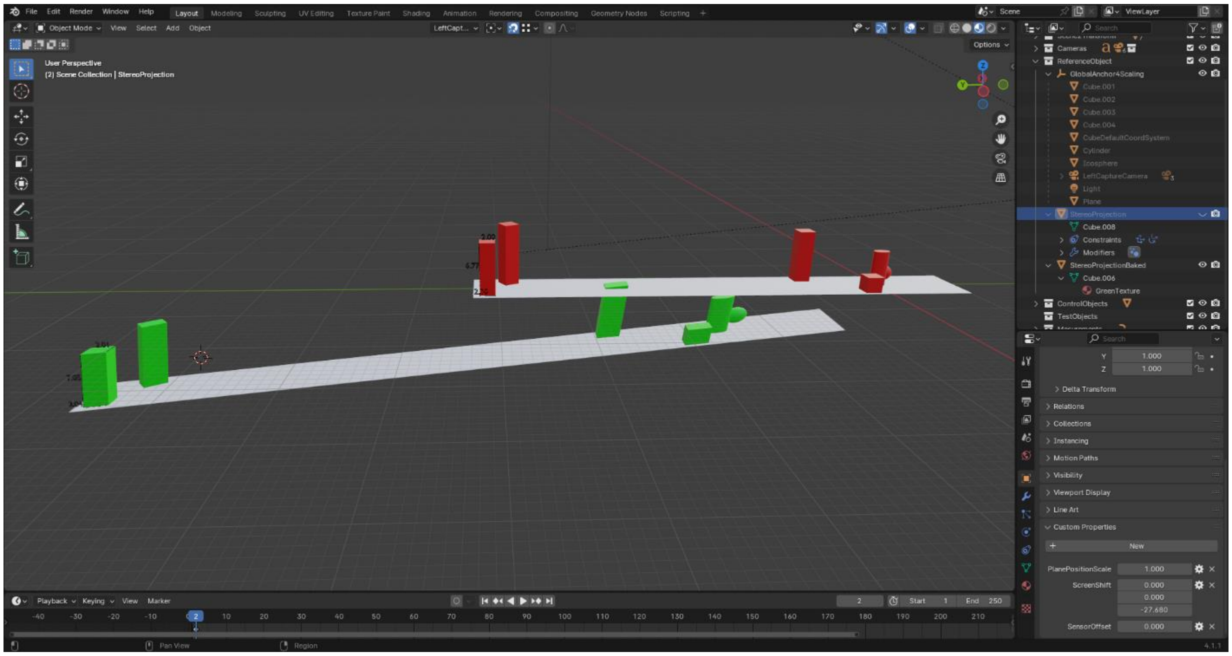


Figure 10: Illustration of the stereo parameters (italic) that can be defined in our rapid prototype type tool.

Each of these parameters can be set in Blender, and our plugin transforms the original remodeled theater geometry into the new geometry which corresponds to the perception by the observer. Figure 11 illustrates a synthetic example scene to visualize the principles. The original scene consists of a floor plane and a couple of cubes and cylinders (red) that are supposed to be recorded by a stereo camera. Then our Blender plugin computes a new 3D model (green objects) corresponding to what a user perceives when the stereo parameters are modified as desired. This approach also allows us to detect illegal stereo parameters causing negative disparity. Such a scenario requires an observer to squint outwards, which is very unnatural and results in eyestrain. The sensor offset and the mismatch between the position of the user's eyes and the recording stereo cameras are critical parameters in this regard.



PlanePositionScale 1.000

ScreenShift 0.000 0.000 -27.680

SensorOffset 0.000

Figure 11: Original example scene (red cubes on a plane) recorded by a stereo camara, and how this scene is perceived when the focal length of the projection is changed (green cubes on a plane).

By these means, we can adjust the parameters in such a way that the perceived stereo impression matches with the 3D-model of the theater. The left setup in Figure 12, for instance, causes a gap between the stereo and the CGI floor, leading to an odd user perception. This can be fixed by moving and rotating the stereo projectors.

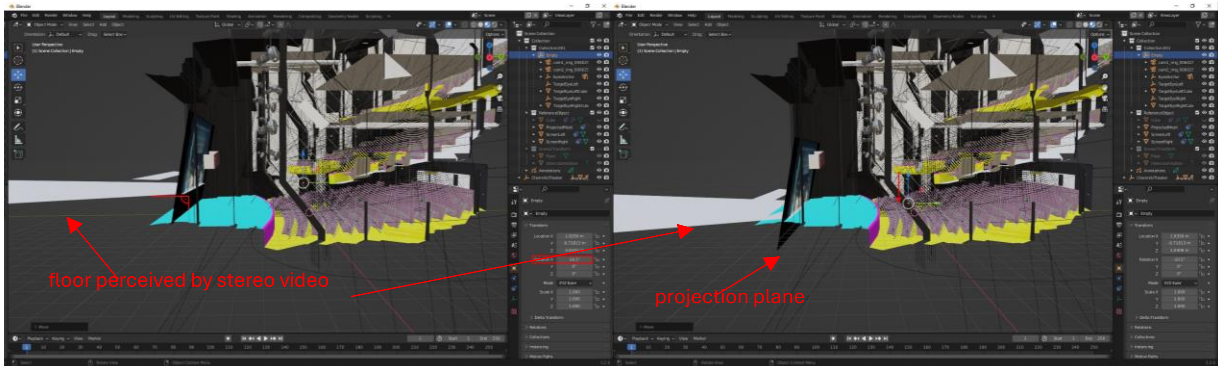


Figure 12: 3D-modell of a theater in combination with a 3D-modell of the floor reconstructed from the stereo video. Left: Wrong stereo parameters cause a mismatch between CGI floor (blue) and the floor recorded by the stereo cameras (gray plane). Right: Smooth transition of the stereo floor and the CGI floor.

Lastly, we need to define the position of the projection planes by moving them in the camera frustum. To avoid inconsistencies when the observer moves his or her head, they should be as close as possible to the transition between the CGI and the stereo objects (see Figure 12 right). Moreover, relevant video content must not be occluded by the 3D-model of the floor or any other element of the event location.

5.3 Unity application

Once the projection screens have been correctly placed relative to the 3D-model of the event location, they can be exported from Blender and imported into Unity, which we use for the overall VR application. The rectified stereo videos are encoded in a single movie file by vertically packing the left and the right image into one frame (see Figure 13). It is then assigned as texture to both projection planes. By adjusting the UV-map of each projection plane, we ensure that the left image is visualized on the left projection plane, and the right image is visualized on the right projection plane. For decoding the video itself, we use standard Unity tools.

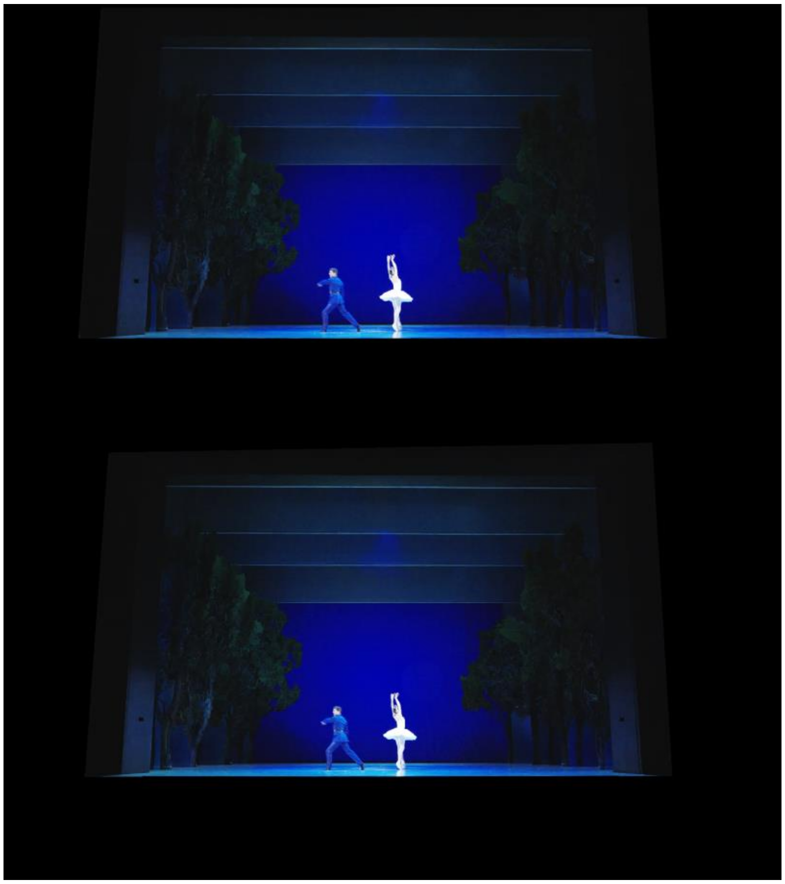


Figure 13: "Over/Under" packing of the left and right images into one stereo video frame.

Due to separate capture of the event location and the live act, we face the challenge of inconsistent illumination. As remedy, we allow a smooth blending between the stereo video and the 3D environment by adding transparency to the bottom parts of the stereo video. We only need to ensure that those video parts do not show the actors themselves, otherwise we mix actor pixels with floor pixels. Moreover, we dynamically relight the theater model based on sample patches from the stereo video.

Finally, we hide seams between the stereo content and the 3D-model of the event location that could not be eliminated by adjusting the stereo parameters. This can for instance happen, when the stereo baseline needs to be corrected too strongly because the camera bodies are too huge for correct inter-eye distance. This results in a scaling of the objects visualized by the stereo video, which is hard to compensate perfectly. To solve this problem, we ensured proper matching of the floor in the 3D-model and the stereo video, while seams at the walls were concealed by additional curtains (see Figure 14).

Figure 14: Actors' live act integrated into 3D model of event location populated by avatars.

6 PROOF OF CONCEPT PRODUCTION AND LESSONS LEARNED

For proof of concept, we have applied the workflow mentioned above in several shows of the Theater in Chemnitz, Germany. They covered pieces such as Swan Lake (72) and Cinderella (73). The recording happened during the dress rehearsals.

From these practical experiences, we can summarize our lessons learned as follows:

- Clarify beforehand which capture location does not interfere with the course of the event
- Ensure that the camera view towards the stage is not blocked/occluded by persons or objects (i.e. conductor)
- Avoid direct incidence of spotlights towards cameras
- Have your stereo camera close to a real seating position and orientation
- Try to have a frontal view on the live act
- Check for flickering lights and adjust your framerates as well as possible
- Use small, synchronized cameras with high sensitivity and high dynamic range
- Synchronize settings between both cameras (focus, white balance, ...)
- Keep baseline as close as possible to the inter-eye distance
- Ensure an all-in-focus capture
- Try to have the optical axis of camera orthogonal to the direction of gravity
- Ensure that all regions where actors could go to are covered by your field of view of both cameras

- Measure dimensions of the stage and camera setup for later reference
- Make enough making-of photos to document the capture setting and the stage setup in good lighting conditions
- Record with high video bitrate

7 CONCLUSION AND OUTLOOK

This paper has presented a novel workflow how to integrate stereo content of a live act into a VR experience. It avoids motion sickness by splitting the capture of the event location and of the actors' performance, so that sufficient parallax is available for close objects. At the same time, it allows to reduce the capture costs by resorting to stereo cameras only for the live act. By these means, we provide a new tradeoff between viewing accuracy and production costs.

We showed that such an approach is possible, when there is a sufficient distance between the actors and the audience. In such cases, the missing parallax of the stereo capture showed to be hardly recognizable, because the event location is reproduced with full parallax and leads to natural viewing experiences. In contrast to mono-video, stereo offers a better and more realistic depth representation. However, compared to traditional stereo movies for cinema or television, a more complex compositing approach is necessary to seamlessly integrate the stereo content into the VR environment.

We have come up with a rigorous method how to achieve this integration with high quality. The necessary steps only depend on the scene geometry and the camera setup. Once those are fixed, we only need to stream a traditional stereo video signal to the viewing clients. We believe that this opens new applications to current broadcasting infrastructure. The capability to reduce latency as today already required for instance in sports transmissions becomes also crucial when there should be an interaction between the actors and the virtual audience. As today's events live from the interaction between the actors and the physical audience, this would also bring virtual attendance to a new level.

8 ACKNOWLEDGEMENTS

The authors gratefully acknowledge the scientific support and HPC resources provided by the Erlangen National High Performance Computing Center (NHR@FAU) of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). The hardware is funded by the German Research Foundation (DFG).

GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

The project on which this publication is based on was funded by the German Federal Ministry of Education and Research under grant numbers 16SV8773 and 16SV8775. Responsibility for the content of this publication lies with the authors.

9 REFERENCES

1. Parallax. In: Wikipedia [Internet]. 2024 [cited 2024 Jul 22]. Available from: https://en.wikipedia.org/w/index.php?title=Parallax&oldid=1221302278

2. VRChat [Internet]. [cited 2024 Aug 1]. Available from: https://hello.vrchat.com/

3. Metallica Rocks Fortnite with a New Music Experience & More! [Internet]. [cited 2024 Aug 1]. Available from: https://www.fortnite.com/news/metallica-lights-up-fortnite-with-a-new-music-experience-festival-season-and-more

4. Winkler A, Won J, Ye Y. QuestSim: Human Motion Tracking from Sparse Sensors with Simulated Avatars. In: SIGGRAPH Asia 2022 Conference Papers [Internet]. Daegu Republic of Korea: ACM; 2022 [cited 2023 Jun 2]. p. 1–8. Available from: https://dl.acm.org/doi/10.1145/3550469.3555411

5. Xia S, Gao L, Lai YK, Yuan MZ, Chai J. A Survey on Human Performance Capture and Animation. J Comput Sci Technol. 2017 May;32(3):536–54.

6. Weissig C, Schreer O, Eisert P, Kauff P. The Ultimate Immersive Experience: Panoramic 3D Video Acquisition. In: Schoeffmann K, Merialdo B, Hauptmann AG, Ngo CW, Andreopoulos Y, Breiteneder C, editors. Advances in Multimedia Modeling [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012 [cited 2023 Jun 1]. p. 671–81. (Lecture Notes in Computer Science; vol. 7131). Available from: http://link.springer.com/10.1007/978-3-642-27355-1_72

7. Matzen K, Cohen MF, Evans B, Kopf J, Szeliski R. Low-cost 360 stereo photography and video capture. ACM Trans Graph. 2017 Jul 20;36(4):1–12.

8. Hannuksela MM, Wang YK, Hourunranta A. An Overview of the OMAF Standard for 360° Video. In: 2019 Data Compression Conference (DCC). 2019. p. 418–27.

9. Sreedhar KK, Curcio IDD, Hourunranta A, Lepistö M. Immersive media experience with MPEG OMAF multi-viewpoints and overlays. In: Proceedings of the 11th ACM Multimedia Systems Conference [Internet]. Istanbul Turkey: ACM; 2020 [cited 2023 Jul 9]. p. 333–6. Available from: https://dl.acm.org/doi/10.1145/3339825.3393576

10. A rapid prototyping tool to produce 360° video-based immersive experiences enhanced with virtual/multimedia elements. Procedia Computer Science. 2018 Jan 1;138:441–53.

11. Wang M, Lyu XQ, Li YJ, Zhang FL. VR content creation and exploration with deep learning: A survey. Comp Visual Media. 2020 Mar;6(1):3–28.

12. Valenzise G, Alain M, Zerman E, Ozcinar C, editors. Immersive Video Technologies [Internet]. Academic Press; 2022. 630 p. Available from: https://www.elsevier.com/books/immersive-video-technologies/valenzise/978-0-323-91755-1

13. Perazzi F, Sorkine-Hornung A, Zimmer H, Kaufmann P, Wang O, Watson S, et al. Panoramic Video from Unstructured Camera Arrays. Computer Graphics Forum. 2015 May;34(2):57–68.

14. Complete List of VR Camera Systems 2019 [Internet]. 2019 [cited 2020 Jul 15]. Available from: https://delight-vr.com/blog/complete-list-of-vr-cameras-2019/

15. Schroers C, Bazin JC, Sorkine-Hornung A. An Omnistereoscopic Video Pipeline for Capture and Display of Real-World VR. ACM Trans Graph. 2018 Jun 30;37(3):1–13.

16. User I. Introducing Facebook Surround 360: An open, high-quality 3D-360 video capture system [Internet]. Engineering at Meta. 2016 [cited 2023 Jul 8]. Available from: https://engineering.fb.com/2016/04/12/video-engineering/introducing-facebook-surround-360-an-open-high-quality-3d-360-video-capture-system/

17. Anderson R, Gallup D, Barron JT, Kontkanen J, Snavely N, Hernández C, et al. Jump: virtual reality video. ACM Trans Graph. 2016 Nov 11;35(6):1–13.

18. Attal B, Ling S, Gokaslan A, Richardt C, Tompkin J. MatryODShka: Real-time 6DoF Video View Synthesis using Multi-Sphere Images. arXiv:200806534 [cs] [Internet]. 2020 Aug 14 [cited 2021 May 3]; Available from: http://arxiv.org/abs/2008.06534

19. Dziembowski A, Mieloch D, Stankiewicz O, Domański M, Lee G, Seo J. Virtual View Synthesis for 3DoF+ Video. In: 2019 Picture Coding Symposium (PCS). 2019. p. 1–5.

20. Dashwood T. A Beginner’s Guide to Shooting Stereoscopic 3D.

21. Wimmer P. Aufnahme und Wiedergabe stereoskopischer Videos.

22. Xu D, Coria L, Nasiopoulos P. Guidelines for capturing high quality stereoscopic content based on a systematic subjective evaluation. In: 2010 17th IEEE International Conference on Electronics, Circuits and Systems [Internet]. Athens, Greece: IEEE; 2010 [cited 2023 Jul 8]. p. 162–5. Available from: http://ieeexplore.ieee.org/document/5724479/

23. Gunnewiek RK, Vandewalle P. HOW TO DISPLAY 3D CONTENT REALISTICALLY.

24. Lang M, Hornung A, Wang O, Poulakos S, Smolic A, Gross M. Nonlinear disparity mapping for stereoscopic 3D. ACM Trans Graph. 2010 Jul 26;29(4):1–10.

25. Hornung A, Smolic A, Gross M. Novel Stereoscopic Content Production Tools.

26. Zilly F, Kluger J, Kauff P. Production Rules for Stereo Acquisition. Proc IEEE. 2011 Apr;99(4):590–606.

27. Zhang J, Liu X, Ye X, Zhao F, Zhang Y, Wu M, et al. Editable Free-viewpoint Video Using a Layered Neural Representation. arXiv:210414786 [cs] [Internet]. 2021 Apr 30 [cited 2021 May 3]; Available from: http://arxiv.org/abs/2104.14786

28. Collet A, Chuang M, Sweeney P, Gillett D, Evseev D, Calabrese D, et al. High-quality streamable free-viewpoint video. ACM Trans Graph. 2015 Jul 27;34(4):1–13.

29. Martin-Brualla R, Pandey R, Yang S, Pidlypenskyi P, Taylor J, Valentin J, et al. LookinGood: Enhancing Performance Capture with Real-time Neural Re-Rendering. arXiv:181105029 [cs] [Internet]. 2018 Nov 12 [cited 2020 Aug 6]; Available from: http://arxiv.org/abs/1811.05029

30. Choi J, Jeong JB, Lee S, Ryu ES. Overview of the Volumetric Video Capturing System for Immersive Media. In: 2022 13th International Conference on Information and Communication Technology Convergence (ICTC) [Internet]. Jeju Island, Korea, Republic of: IEEE; 2022 [cited 2023 Jul 8]. p. 574–7. Available from: https://ieeexplore.ieee.org/document/9953020/

31. Joo H, Liu H, Tan L, Gui L, Nabbe B, Matthews I, et al. Panoptic Studio: A Massively Multiview System for Social Motion Capture. In: 2015 IEEE International Conference on Computer Vision (ICCV) [Internet]. Santiago, Chile: IEEE; 2015 [cited 2023 Jul 8]. p. 3334–42. Available from: http://ieeexplore.ieee.org/document/7410738/

32. de Aguiar E, Stoll C, Theobalt C, Ahmed N, Seidel HP, Thrun S. Performance Capture from Sparse Multi-view Video.

33. Alldieck T, Magnor M, Xu W, Theobalt C, Pons-Moll G. Video Based Reconstruction of 3D People Models. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition [Internet]. Salt Lake City, UT: IEEE; 2018 [cited 2023 Jul 8]. p. 8387–97. Available from: https://ieeexplore.ieee.org/document/8578973/

34. Martínez-Cano FJ. Volumetric filmmaking, new mediums and formats for digital audiovisual storytelling. A|C [Internet]. 2021 Feb 26 [cited 2023 Jul 8]; Available from: https://publication.avanca.org/index.php/avancacinema/article/view/168

35. Eisert P, Schreer O, Feldmann I, Hellge C, Hilsmann A. Volumetric video – acquisition, interaction, streaming and rendering. In: Valenzise G, Alain M, Zerman E, Ozcinar C, editors. Immersive Video Technologies [Internet]. Academic Press; 2023 [cited 2023 May 25]. p. 289–326. Available from: https://www.sciencedirect.com/science/article/pii/B9780323917551000171

36. Wu G, Masia B, Jarabo A, Zhang Y, Wang L, Dai Q, et al. Light Field Image Processing: An Overview. IEEE J Sel Top Signal Process. 2017 Oct;11(7):926–54.

37. Herfet T. 5D Light Field Video Capture.

38. Sabater N, Boisson G, Vandame B, Kerbiriou P, Babon F, Hog M, et al. Dataset and Pipeline for Multi-view Light-Field Video. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2017. p. 1743–53.

39. Flynn J, Broxton M, Debevec P, DuVall M, Graham Fyffe, Overbeck R, et al. DeepView: View Synthesis With Learned Gradient Descent. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. Long Beach, CA, USA: IEEE; 2019 [cited 2020 May 12]. p. 2362–71. Available from: https://ieeexplore.ieee.org/document/8953705/

40. Pan Y, Liu Y, Zhang L. LiTrix: A Lightweight Live Light Field Video Scheme for Metaverse Stereoscopic Applications. IEEE Internet Things M. 2023 Jun;6(2):137–42.

41. Ziegler M, Engelhardt A, Müller S, Keinert J, Zilly F, Foessel S, et al. Multi-camera system for depth based visual effects and compositing. In New York, NY, USA: Association for Computing Machinery; 2015 [cited 2021 Mar 23]. p. 1–10. (CVMP ’15). Available from: https://doi.org/10.1145/2824840.2824845

42. Yang JC, Everett M, Buehler C, McMillan L. A Real-Time Distributed Light Field Camera. :10.

43. Herfet T, Chelli K, Le Pendu M. Chapter 6 - Acquisition of light field images & videos: Capturing light rays. In: Valenzise G, Alain M, Zerman E, Ozcinar C, editors. Immersive Video Technologies [Internet]. Academic Press; 2023 [cited 2023 Jul 15]. p. 163–71. Available from: https://www.sciencedirect.com/science/article/pii/B9780323917551000122

44. Herfet T, Chelli K, Le Pendu M. Chapter 7 - Light field representation: The dimensions in light fields. In: Valenzise G, Alain M, Zerman E, Ozcinar C, editors. Immersive Video Technologies [Internet]. Academic Press; 2023 [cited 2023 Jul 15]. p. 173–99. Available from: https://www.sciencedirect.com/science/article/pii/B9780323917551000134

45. Feng BY, Jabbireddy S, Varshney A. VIINTER: View Interpolation with Implicit Neural Representations of Images. In: SIGGRAPH Asia 2022 Conference Papers [Internet]. Daegu Republic of Korea: ACM; 2022 [cited 2022 Dec 7]. p. 1–9. Available from: https://dl.acm.org/doi/10.1145/3550469.3555417

46. Shum H, Kang SB. Review of image-based rendering techniques. In: Visual communications and image processing 2000. 2000. p. 2–13.

47. Levoy M, Hanrahan P. Light field rendering. In: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques - SIGGRAPH '96 [Internet]. Not Known: ACM Press; 1996 [cited 2020 Jun 30]. p. 31–42. Available from: http://portal.acm.org/citation.cfm?doid=237170.237199

48. Buehler C, Bosse M, McMillan L, Gortler S, Cohen M. Unstructured lumigraph rendering. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques (SIGGRAPH) [Internet]. Los Angeles, CA, USA: ACM Press; 2001 [cited 2020 Jun 6]. p. 425–32. Available from: http://portal.acm.org/citation.cfm?doid=383259.383309

49. Gyeong Ja Jang, Keun Ho Kim, In Kyu Park. Depth image-based rendering of 3D object on mobile device. In: Proceedings of 2004 International Symposium on Intelligent Signal Processing and Communication Systems, 2004 ISPACS 2004. 2004. p. 630–3.

50. Bogaert LV, Bonatto D, Fachada S, Lafruit G. Novel view synthesis in embedded virtual reality devices. The Engineering Reality of Virtual Reality. 2022;6.

51. Penner E, Zhang L. Soft 3D reconstruction for view synthesis. ACM Trans Graph. 2017 Nov 20;36(6):1–11.

52. Lange T, Petrovic G, Herfet T. Real-time virtual view rendering for video communication systems using OpenCL. In: 2014 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting. 2014. p. 1–5.

53. Vandame B, Sabater N, Boisson G, Doyen D, Allié V, Babon F, et al. Pipeline for Real-Time Video View Synthesis. In London, UK; 2020. p. 1–6.

54. Ziegler M. Advanced image processing for immersive media applications using sparse light-fields [PhD Thesis]. [Erlangen]: Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU); 2019.

55. Broxton M, Flynn J, Overbeck R, Erickson D, Hedman P, DuVall M, et al. Immersive Light Field Video with a Layered Mesh Representation. ACM Transactions on Graphics. 2020 Jul;39(4):15.

56. Mildenhall B, Srinivasan PP, Ortiz-Cayon R, Kalantari NK, Ramamoorthi R, Ng R, et al. Local Light Field Fusion: Practical View Synthesis with Prescriptive Sampling Guidelines. arXiv:190500889 [cs] [Internet]. 2019 May 2 [cited 2020 May 12]; Available from: http://arxiv.org/abs/1905.00889

57. Gao K, Gao Y, He H, Lu D, Xu L, Li J. NeRF: Neural Radiance Field in 3D Vision, A Comprehensive Review [Internet]. arXiv; 2022 [cited 2022 Oct 10]. Available from: http://arxiv.org/abs/2210.00379

58. Mildenhall B, Srinivasan PP, Tancik M, Barron JT, Ramamoorthi R, Ng R. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In: Vedaldi A, Bischof H, Brox T, Frahm JM, editors. Cham: Springer; 2020. p. 405–21.

59. Xie Y, Takikawa T, Saito S, Litany O, Yan S, Khan N, et al. Neural Fields in Visual Computing and Beyond. 2022 Apr 5 [cited 2022 Jun 22]; Available from: http://arxiv.org/abs/2111.11426

60. Virtualization Chemnitz Opera House Part 1: Photogrammetry – SocialSTAGE-VR [Internet]. [cited 2024 Jul 8]. Available from: https://socialstagevr.de/virtualization-chemnitz-opera-house-part-1-photogrammetry/

61. Virtualization Chemnitz Opera House Part 2: Post-Modeling and Optimization – SocialSTAGE-VR [Internet]. [cited 2024 Jul 8]. Available from: https://socialstagevr.de/virtualization-chemnitz-opera-house-part-2-post-modeling-and-optimization/

62. Ziegler M, Keinert J, Holzer N, Wolf T. IMMERSIVE VIRTUAL REALITY FOR LIVE-ACTION VIDEO USING CAMERA ARRAYS.

63. RX0 II premium tiny tough camera | DSC-RX0M2G | Sony United Kingdom [Internet]. [cited 2023 Jul 29]. Available from: https://www.sony.co.uk/electronics/cyber-shot-compact-cameras/dsc-rx0m2?locale=en_GB

64. 3D rig. In: Wikipedia [Internet]. 2023 [cited 2024 Jul 8]. Available from: https://en.wikipedia.org/w/index.php?title=3D_rig&oldid=1157862944

65. Reda F, Kontkanen J, Tabellion E, Sun D, Pantofaru C, Curless B. FILM: Frame Interpolation for Large Motion. In: Avidan S, Brostow G, Cissé M, Farinella GM, Hassner T, editors. Computer Vision – ECCV 2022 [Internet]. Cham: Springer Nature Switzerland; 2022 [cited 2024 Jul 8]. p. 250–66. (Lecture Notes in Computer Science; vol. 13667). Available from: https://link.springer.com/10.1007/978-3-031-20071-7_15

66. Blackmagic URSA Cine Immersive [Internet]. [cited 2024 Jul 8]. Available from: https://www.blackmagicdesign.com/media/release/20240611-02

67. Schönberger JL, Frahm JM. Structure-from-Motion Revisited. In Las Vegas, NV, USA; 2016 [cited 2021 Sep 3]. p. 4104–13. Available from: http://ieeexplore.ieee.org/document/7780814/

68. Ma J, Jiang X, Fan A, Jiang J, Yan J. Image Matching from Handcrafted to Deep Features: A Survey. Int J Comput Vis. 2021 Jan;129(1):23–79.

69. Scharstein D, Szeliski R, Zabih R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In Kauai, HI, USA; 2001 [cited 2021 Nov 27]. p. 131–40. Available from: http://ieeexplore.ieee.org/document/988771/

70. Hamzah RA, Ibrahim H. Literature Survey on Stereo Vision Disparity Map Algorithms. Journal of Sensors. 2016;1–23.

71. Laga H, Jospin LV, Boussaid F, Bennamoun M. A Survey on Deep Learning Techniques for Stereo-based Depth Estimation. IEEE Trans Pattern Anal Mach Intell. 2022 Apr 1;44(4):1738–64.

72. *Swan Lake*. In: Wikipedia [Internet]. 2024 [cited 2024 Jul 22]. Available from: https://en.wikipedia.org/w/index.php?title=Swan_Lake&oldid=1235185024

73. Cinderella. In: Wikipedia [Internet]. 2024 [cited 2024 Jul 22]. Available from: https://en.wikipedia.org/w/index.php?title=Cinderella&oldid=1235190617