DEPTH AUGMENTED STEREO PANORAMA FOR CINEMATIC VIRTUAL REALITY WITH FOCUS CUES

Jayant Thatte, Jean-Baptiste Boin, Haricharan Lakshman, Gordon Wetzstein, Bernd Girod

Department of Electrical Engineering, Stanford University

ABSTRACT

Cinematic virtual reality (VR) aims to provide immersive visual experiences of real-world scenes on head-mounted displays. Current cinematic VR systems support stereo cues, but not focus cues that are important for depth perception and comfortable viewing. We propose a new content representation, depth augmented stereo panorama (DASP), which permits generating light fields (LFs) across the observer's pupils, achieving an order of magnitude reduction in data requirements compared to the existing techniques. The LF generation and refocusing capabilities of DASP are evaluated for both, computer-generated and real scenes. Results indicate that DASP can successfully create stereo as well as support focus cues.

***Index Terms*—** Virtual reality, image representation, focus cues

1. INTRODUCTION

With the availability of low cost, wide field-of-view, fast refresh rate head-mounted displays (HMDs), virtual reality (VR) is rapidly gaining popularity. VR has recently gained a lot of applications in entertainment, car and flight simulators [1] and in medicine [2]. Historically, VR has been associated with synthetic content, computer generated using 3D models. However, recent advances in capture devices, for example spherical camera rigs or 360° cameras with wide angle optics, facilitate omnidirectional capture of real-world environments, making it possible to create live-action videos of real-world scenes, to be viewed on HMDs - a technology referred to as cinematic VR.

Studies of human visual perception have shown that our brains perceive depth by simultaneously combining information from several visual cues [3]. Conventional HMDs provide stereo depth cues using binocular disparity. However, they do not support focus cues. As a result, the existing HMDs are known to induce a *vergence-accommodation conflict* (VAC) [4] which contributes to visual discomfort, eye strain and nausea. It has also been reported that retinal blur is a dominant depth cue for scene regions lying far from the vergence plane [5]. Thus, accurate rendering of focus cues is pivotal for a comfortable and immersive experience, which will be a key factor in the acceptance of VR as a mainstream technology. Realizing VAC as a significant challenge in VR, there's been a lot of work on the development of new types of VR HMDs that attempt to solve this problem - for example using microlens arrays [6], stacked LCD-type LF displays [7], multiplane displays [8, 9, 10, 11], holographic displays, and so on. A natural framework to support focus cues in cinematic VR would be to represent the VR content as a LF consisting of several sub-aperture images over the pupillary area. Studies have shown that a collection of such sub-aperture images can be used to achieve a plausible approximation of retinal blur [12]. However, this naïve approach would result in prohibitively large data requirements to support omnidirectional focus cues. Furthermore, omnidirectional stereoscopic LF capture and editing for dynamic scenes is challenging.

Format	Stereo in all directions	Dis-occlusion handling	Focus cues for all azimuth	Focus cues for all elevation
Mono Pano.	✗	✗	✗	✗
Mono Pano. + Depth	✓	✗	✓	✓
Omnistereo [13]	✓	✓	✗	✗
Concentric Mosaics [14]	✓	✓	✓	✗
Central LF Pano. [15]	✗	✗	✓	✗
DASP	✓	✓	✓	✓

Table 1: A summary of 3D video formats for supporting stereo and focus cues in cinematic VR application.

We propose a novel data representation called depth augmented stereo panorama (DASP) for cinematic VR that supports both stereo and focus cues. We show that DASP has an order of magnitude smaller data requirements compared to the existing methods and is able to produce high quality content for cinematic VR. In addition, the proposed representation is flexible enough to be compatible with different HMD technologies and fits well into the existing content creation and distribution infrastructures. It is also able to support variations in pupil size and inter-pupillary distance (IPD).

2. RELATED WORK

Formats with explicit geometry offer immense possibilities during rendering. However, most real-world scenes are too complex and rich in details for reconstructing high quality 3D models. On the other end of the spectrum, monoscopic panoramic representations are compact but do not support disparity between the left and right views. Auxiliary depth information could, in theory, be used to generate stereo panoramas. However, such a format cannot handle dis-occlusions resulting from perspective changes due to head rotation. A summary of 3D video formats in terms of omnidirectional stereo and focus cues capabilities is provided in Table 1.

Omnistereo panoramas: No fixed arrangement of two single-perspective 360° cameras can provide stereo in all directions, since there would be no parallax along the extended baseline. To tackle this, Peleg *et al.* [13] proposed *omnistereo*, a multi-perspective panoramic representation that can support stereo vision in all directions. Techniques such as [16] can be used to mitigate the stitching seams when using a limited number of cameras. However, omnistereo representations are not capable of synthesizing LFs required for supporting focus cues. Moreover, they only model head rotation in equatorial plane and do not support looking up or down in a scene.

Concentric mosaics: Shum *et al.* [14] proposed concentric mosaics as an extension of omnistereo, where a set of omnistereo panoramas at different viewing radii are stored. This can be used to generate LFs with horizontal perspective shifts, but not a 2D set of perspectives required for proper focus cues in both horizontal and vertical directions. Furthermore, since ray space interpolation using im-

ages from different perspectives may lead to vertical distortions, these have to be corrected with depth information [14]. Therefore, although intended to be purely image-based, depth information is additionally required in practical applications. Since several omnistereo images have to be captured to synthesize high quality novel views, the data overhead of using this representation is significant.

Cylindrical LF panoramas: Birklbauer *et al.* [15, 17] proposed a cylindrical parameterization of LF panoramas that can be used for refocusing a monoscopic panorama. However, it would only be able to render focus cues for viewing directions in the equatorial plane. When a viewing direction has a non-zero elevation angle, the appropriate pupil plane is no longer vertical, but is tilted by the corresponding angle. For example, the pupil plane is completely horizontal when a viewer is looking directly up or down. Thus, to be able to produce focus cues outside of equatorial viewing directions, a LF with a set of perspectives on a non-vertical viewing plane would be required. Moreover, the LF data in this representation might have to be oversampled to counter potential aliasing artifacts in rendered viewports when device specifications or user requirements vary.

3. DEPTH AUGMENTED STEREO PANORAMAS

3.1. Proposed Omnistereo Texture-plus-depth Format

Considering the pros and cons of existing representations and the requirement to support truly omnidirectional stereoscopic LFs, we propose *depth augmented stereo panorama* (DASP). Since a rotating camera is not suitable for acquiring dynamic scenes, constructions such as camera rigs would be needed to simultaneously capture rays in different directions. Additionally, the capture system necessarily has to be multi-perspective to be able to generate stereo in all directions. Therefore, computational techniques are employed to stitch together panoramic representations of the scene. To achieve seamless stitching in spite of parallax between captured views, correspondence/depth estimation is typically carried out in the postproduction stage. Instead of discarding the depth information after stitching, we propose to remap the depth information to the same omnistereo format as the textures and store the omnistereo texture-plus-depth panoramas.

Spherical Omnistereo: The classical formulation of omnistereo has a limited vertical field of view. The points in the world that lie directly above or below the area of the viewing circle do not contribute to the panoramas and hence cannot be captured and represented [13]. Since VR applications require a full 180° vertical and 360° horizontal view, it becomes critical that any representation used in VR be

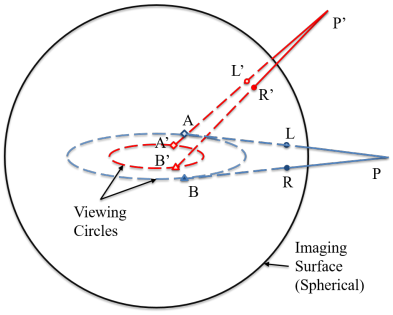

Fig. 1: Spherical Omnistereo. P and P' represent equatorial and non-equatorial points, respectively. For the point P, A and B are the corresponding eye-positions, while for P' they are A' and B'. Notice how the viewing radius shrinks as the scene point elevation changes.

able to represent the whole space. With this in mind, we propose an extension to omnistereo. We modify the imaging surface from cylindrical to spherical and also use a viewing "disk" rather than a viewing circle. For points on the equatorial plane, our representation is identical to the classical omnistereo. However, to map points at a non-zero elevation, we use a smaller viewing radius, by a factor of cosine of the elevation angle. Thus, scene points at different elevations are mapped onto different viewing circles, which collectively form the viewing disk. This allows us to represent the entire 3D space, unlike classical omnistereo.

This means that considering the image information alone, the angular disparity for a given depth reduces and finally becomes zero as we look completely up or down towards the poles [18]. However, in our proposed formulation, we additionally have depth information and therefore we can create stereo for all elevations from -90° to 90°. LF synthesis from DASP is performed in two stages: scene reconstruction, followed by LF creation by forward projection onto the desired viewpoints.

3.2. Scene Reconstruction using Proposed Format

The goal is to reconstruct a point cloud by mapping every pixel (h,w) in each panorama in DASP to a scene point P, represented by a 3D vector $\overrightarrow{rP}$. Let O be the center of the coordinate system, which is also the center of the imaging sphere and the viewing disk. Let f and v be the radii of the imaging sphere and the viewing disk, respectively. For a pixel (h,w), let $L=(f,\theta_L,\phi_L)$ and $A=(v\cos\phi_L,\theta_A,0)$ be the corresponding image point and eye position respectively (referring to Figure 1), where $[\theta_L,\phi_L]=\mathscr{L}(h,w)$, $\mathscr{L}$ is the linear mapping from equirectangular pixel coordinates to polar image coordinates and the $\cos\phi_L$ term is to capture the shrinking viewing radius with elevation angle. By construction, $\overrightarrow{OA}\cdot\overrightarrow{AL}=0$. This gives $\theta_A=\theta_L\pm\cos^{-1}(v/f)$, positive for left eye. Additionally, A, L and P are collinear, and $|\overrightarrow{AP}|=d$, where d is the augmented depth value corresponding to pixel (h,w). Thus, $\overrightarrow{rP}=\overrightarrow{OA}+\overrightarrow{AP}=\overrightarrow{OA}+(d/|\overrightarrow{AL}|)\overrightarrow{AL}$, which can be computed since we know d and coordinates of A and L. Thus, the coordinates of P can be determined, knowing (h,w,d).

3.3. Light Field Synthesis

For generating light fields, we use an eye box of 1 cm × 1 cm in order to address different pupil positions and sizes [19].

1. *Viewpoints definition*: Given a viewing direction $\hat{r}=(1,\theta,\phi)$ and the radius of the viewing circle v, the location of the pupil center is given by $A=(v\cos\phi,\theta\pm\cos^{-1}(v/f),0)$, positive for left eye, as derived in Sec. 3.2. An $N\times N$ grid is defined centered at the pupil, with pupil plane normal along $\hat{r}$. Target image plane is chosen perpendicular to the viewing direction.
2. *Depth warping*: The point cloud computed in Sec. 3.2, is forward projected onto the target image plane with the desired viewpoint as the center of projection, using half-pixel warping and bilinear interpolation. The disocclusion holes in the warped depth are left unaltered.
3. *Texture warping*: Using the warped depth from above, texture on the target image plane is determined by a lookup into omnistereo texture panoramas generated in Sec. 3.1. This follows similar steps as depth warping, but in the reverse order and uses a bicubic interpolation.
4. *Hypothesis merging*: For each viewpoint, two warped texture hypotheses are generated: one from each of left and right DASP panoramas. These two hypotheses are then merged. If one of the hypothesis has a hole, it's filled using the other. If

both hypotheses have valid textures, then the hypothesis with the smaller corresponding depth value takes precedence. Majority of the holes get filled in the hypothesis merging step.

5. *Hole-filling*: For each of the small fraction of holes that remain, a vector pointing in the general direction of background texture is identified using depth information surrounding the hole. The background pixel values are then propagated into hole areas. Since the holes left after hypothesis merging are usually small in their spatial extent, the results produced by this simple pixel propagation scheme appear plausible.

The generated LFs would then be fed into a light field HMD (LF-HMD). Note that even though we show refocused scenes in the next section for evaluation purposes, in a LF-HMD the refocusing would happen optically through accommodation mechanism of the human visual system.

3.4. Dimensionality Reduction Analysis

An $N \times N$ LF panorama stores N^2 panoramas per eye, as opposed to DASP, which needs only 2 (texture and depth). Thus, DASP would achieve a dimensionality reduction of 12.5 for a 5×5 LF, for instance. Another approach would be to store a focal stack. Assuming an overall depth range of 6 diopters, in steps of ± 0.3 diopter [20], one would need about 18 panoramas, which is 9 times higher than the proposed representation. Thus, DASP achieves a dimensionality reduction of about one order of magnitude compared to both LF panorama and focal stack approaches, while still maintaining the capability to accurately produce omnidirectional stereo and focus cues.

4. RESULTS

The proposed approach is evaluated with both synthetic and real-world data. The experiments with synthetic data are used to validate and quantify the functionality of the proposed approach by comparing it with viewports created using full access to geometry.

4.1. Synthetic Scene Setup

Synthetic scenes and test inputs in the proposed DASP format are generated using *Blender*[1], an open-source 3D computer graphics software which provides a photo-realistic ray-tracer. Our test scene consists of textured objects at various distances from the camera, from 25 cm to 10 m. The background is a hemispherical sky texture at optical infinity, while the floor is textured with a checkerboard pattern where each square has a side of 1 m.

The eye positions are constrained to a viewing circle of 60 mm diameter, corresponding to a typical IPD. We set up an array of 5×5 pinhole cameras over a square of side 10 mm around a given eye position to create our ground truth LF for the eye box. This is used to synthetically refocus at a set of focal distances to obtain ground truth for refocus tests. The proposed input panoramas are generated for a viewing circle of diameter 70 mm to account for 5 mm disparity of the outermost LF view compared to the central view.

4.2. Light Field Comparison

In this experiment, we compare the LFs synthesized by the proposed representation with the ground truth from Blender, for one eye, using 10 selected viewing directions on a sphere, spanning different values of azimuth and elevation angles. For each LF, the 5×5 constituent

[1] https://www.blender.org

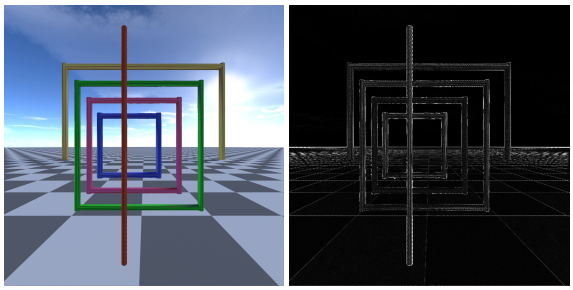

Fig. 2: Left: image synthesized by the proposed representation; Right: pixel-wise absolute difference compared to the ground truth, amplified by a factor of 8. The ground truth image is not shown since the differences between our results and the ground truth are subjectively imperceptible. It can be seen from the amplified error maps that small errors are introduced at depth discontinuities.

images are compared pairwise using PSNR, as well as SSIM [21]. Figure 2 depicts a resulting image and its error compared to the ground truth. The average PSNR over $10 \times 5 \times 5$ images is observed to be 32.9 dB and the corresponding average SSIM is 0.988, with standard deviations of 0.8 dB and 0.002 respectively. These results indicate that the generated LFs approximate the ground truth closely and that the quality is consistent across the synthesized images.

On comparing LFs from DASP with those from monoscopic texture-plus-depth panoramas, we notice that the main differences arise in regions with disocclusions, where the monoscopic representation introduces inpainting artifacts, while the proposed representation does not, since most of the disocclusion holes after warping the left input panorama can be filled by warping the right input panorama, and vice versa. For example, for the viewing direction depicted in the Figure 2, the stereo and mono panoramas achieve PSNR of 34.6 dB and 30.0 dB respectively, and corresponding SSIM of 0.991 and 0.982 respectively.

4.3. Refocus Comparison

Using the generated LFs, we synthetically refocus to simulate an image captured with a larger aperture and choose the plane in focus and compute the PSNR between the refocused images and the ground truth. Note that with a light-field display, refocusing would happen optically through the eye's accommodation. Figure 3 shows an example of refocused viewports for near and far focal distance and Figure 4 shows the quantitative results averaged across several viewing directions. We notice that the PSNR values in the refocus tests are generally higher than the direct LF tests of Sec. 4.2. The

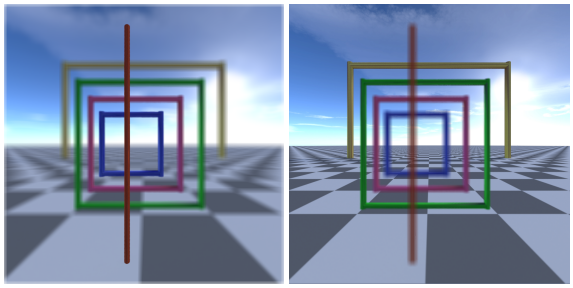

Fig. 3: Example of refocused viewports in a synthetic scene. Left: focus at 25 cm, front brown structure; Right: focus at optical infinity, rear yellow structure and sky.

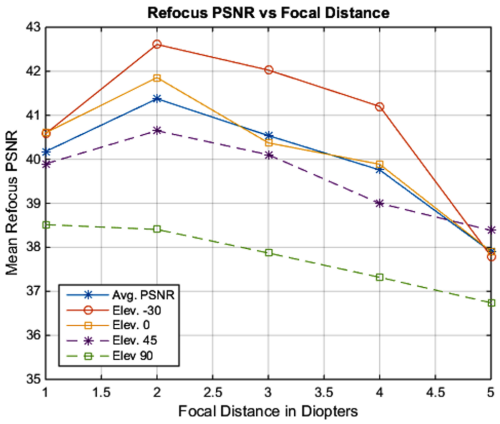

Fig. 4: PSNR of refocused viewports as a function of the focal distance. Different lines in the plot correspond to different elevation angles. For each elevation angle, PSNR of refocused viewports along several azimuth directions were averaged.

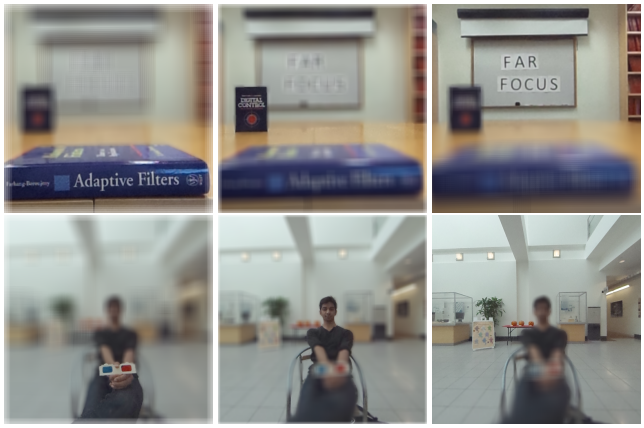

Fig. 5: Examples of refocus on real world scenes. Top, middle and bottom correspond to near, mid and far focus, respectively. In the left column, it can be noticed that exactly one of the texts is in focus in each of the three viewports. In the right column, the near focus is on the anaglyph glasses, mid focus is on the person's face, and far focus is on the rear wall.

main reason for this is that some of the artifacts present at a certain distance are blurred out when refocusing at a different distance. Additionally, the PSNR drops as the absolute elevation angle approaches 90°. This is due to an increase in disocclusion errors near the coordinate poles, which is expected because we reduce the radius of the viewing circle when going towards the poles (as described in Sec. 3.1). The PSNR values for the $-30°$ elevation angle are higher than those of 0° due to the fact that the viewport contains mainly the ground plane at $-30°$ elevation. Across 10 viewing directions and 5 focal distances, the average refocus PSNR is 40 dB with a standard deviation of 1.6 dB, while the average SSIM is 0.997 with a standard deviation of 0.0015.

4.4. Results on Real-world Scenes

For real-world scenes, most existing approaches for omnistereo capture use rotating cameras, which are not suitable for dynamic environments. Ideally, a camera rig would be required to capture rays in all directions such that it supports the desired viewing space. After the capture, several research and engineering challenges have to be addressed to minimize artifacts in stereo panorama stitching [16]. Since the main focus of this paper is the representation of 3D scene and not the production itself, we build a simple proof-of-concept prototype to show the capabilities of the DASP format. We use a setup with two commercial off-the-shelf point 360° cameras (Ricoh Theta m15, image resolution: 3584×1792) placed one above the other (vertical baseline). This arrangement is preferred since stereo in the vicinity of the equator is more important than that for directions looking up or down. Following stereo matching in equirectangular format, omnidirectional depth values are computed using triangulation [22]. Finally, errors in the texture and depth panoramas can be corrected during postproduction, together with creative depth manipulations for a compelling visual experience. A visualization of refocused images using synthesized LFs is shown in Figure 5. For these results, we used LFs of size 9×9, which is enough to mitigate sampling artifacts. Another interesting observation is that despite the depth map obtained with our setup being far from perfect, the output images still produce a convincing depth-of-field effect with correct blur around occlusion boundaries.

5. DISCUSSION

In this paper, we propose depth augmented stereo panorama (DASP) as a novel representation for cinematic VR content. DASP is compatible with existing stereo panoramic framework and produces high-quality LFs for omnidirectional focus cues, with only a small memory footprint. We thus demonstrate that a complete 4D LF capture is not necessary to render the limited parallax required to support focus cues.

Although DASP is designed with LF-HMDs in mind, it is backward compatible with the conventional stereoscopic HMDs. Furthermore, the depth panoramas, if unused, do not add a significant data overhead in the overall system since compressed depth panoramas contribute much less to the data requirement, as compared to compressed texture panoramas.

In addition to supporting focus cues, this format can be used to support parallax for head translations, which is important to make the viewing experience more natural and immersive. In order to update content according to head translations, the same representation can be used, but now with a large enough DASP viewing radius so that the viewpoints for translated novel views still lie within the representation radius. This will ensure minimal inpainting requirement and a high quality of synthesized viewports.

The main source of errors in using DASP for cinematic VR is in the depth estimation stage. In the cinematic VR workflow, these errors need to be fixed during postproduction.

6. CONCLUSION

Cinematic VR requires the use of HMDs over extended time periods, and therefore a comfortable, immersive viewing experience is critical. Incorrect focus cues are an important factor that can cause discomfort. Having recognized this problem, researchers are working on HMDs based on LF technology. However, capture, editing and transmission of LFs can be challenging because of the large volumes of data involved. In this paper, we propose a flexible representation for cinematic VR, suitable for emerging light field HMDs. Our proposed representation helps reduce the dimensionality by an order of magnitude compared to a existing LF or focal stack representations. Furthermore, this fits well in existing creation and distribution infrastructures, and is backward compatible to stereoscopic HMDs. We believe this representation to be pivotal in establishing a practical link between emerging plenoptic capture and display systems.

7. REFERENCES

[1] T. Aslandere, D. Dreyer, and F. Pankratz, "Virtual hand-button interaction in a generic virtual reality flight simulator," in *Aerospace Conference, 2015 IEEE*, March 2015, pp. 1–8. 1

[2] A. Ro, A. Hartholt, M. Grimani, A. Leeds, and M. Liewer, "Virtual reality exposure therapy for combat-related posttraumatic stress disorder," *Computer*, vol. 47, no. 7, pp. 31–37, July 2014. 1

[3] James E. Cutting and Peter M. Vishton, "Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth.," in *Perception of Space and Motion*, William Epstein and Sheena Rogers, Eds., Handbook of Perception and Cognition, pp. 69–117. Academic Press, San Diego, 1995. 1

[4] David M Hoffman, Ahna R Girshick, Kurt Akeley, and Martin S Banks, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of vision*, vol. 8, no. 3, pp. 33, 2008. 1

[5] Robert T Held, Emily A Cooper, and Martin S Banks, "Blur and disparity are complementary cues to depth," *Current Biology*, vol. 22, no. 5, pp. 426–431, 2012. 1

[6] Douglas Lanman and David Luebke, "Near-eye light field displays," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6, pp. 220, 2013. 1

[7] Fu-Chung Huang, Kevin Chen, and Gordon Wetzstein, "The light field stereoscope: Immersive computer graphics via factored near-eye light field displays with focus cues," *ACM Trans. Graph. (SIGGRAPH)*, vol. 34, no. 4, pp. 60:1–60:12, 2015. 1

[8] Sheng Liu, Dewen Cheng, and Hong Hua, "An optical see-through head mounted display with addressable focal planes," in *Mixed and Augmented Reality, 2008. ISMAR 2008. 7th IEEE/ACM International Symposium on*. IEEE, 2008, pp. 33–42. 1

[9] Patrick Llull, Noah Bedard, Wanmin Wu, Ivana Tosic, Kathrin Berkner, and Nikhil Balram, "Design and optimization of a near-eye multifocal display system for augmented reality," in *Propagation through and Characterization of Distributed Volume Turbulence and Atmospheric Phenomena*. Optical Society of America, 2015, pp. JTH3A–5. 1

[10] Gordon D Love, David M Hoffman, Philip JW Hands, James Gao, Andrew K Kirby, and Martin S Banks, "High-speed switchable lens enables the development of a volumetric stereoscopic display," *Optics express*, vol. 17, no. 18, pp. 15716–15725, 2009. 1

[11] Jannick P Rolland, Myron W Krueger, and Alexei Goon, "Multifocal planes head-mounted displays," *Applied Optics*, vol. 39, no. 19, pp. 3209–3215, 2000. 1

[12] Fu-Chung Huang, Gordon Wetzstein, Brian A Barsky, and Ramesh Raskar, "Eyeglasses-free display: towards correcting visual aberrations with computational light field displays," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 4, pp. 59, 2014. 1

[13] Shmuel Peleg, Moshe Ben-Ezra, and Yael Pritch, "Omnistereo: Panoramic stereo imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 279–290, 2001. 1, 2

[14] Heung-Yeung Shum and Li-Wei He, "Rendering with concentric mosaics," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1999, pp. 299–306. 1, 2

[15] Clemens Birklbauer and Oliver Bimber, "Panorama light-field imaging," in *Computer Graphics Forum*. Wiley Online Library, 2014, vol. 33, pp. 43–52. 1, 2

[16] Christian Richardt, Yael Pritch, Henrik Zimmer, and Alexander Sorkine-Hornung, "Megastereo: Constructing high-resolution stereo panoramas," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013*. IEEE, 2013, pp. 1256–1263. 1, 4

[17] Clemens Birklbauer, Simon Opelt, and Oliver Bimber, "Rendering gigaray light fields," in *Computer Graphics Forum*. Wiley Online Library, 2013, vol. 32, pp. 469–478. 2

[18] Murray Eisenberg and Robert Guy, "A proof of the hairy ball theorem," *American Mathematical Monthly*, pp. 571–574, 1979. 2

[19] Bernard Kress and Thad Starner, "A review of head-mounted displays (hmd) technologies and applications for consumer electronics," in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2013, pp. 87200A–87200A. 2

[20] Kevin J MacKenzie, David M Hoffman, and Simon J Watt, "Accommodation to multiple-focal-plane displays: Implications for improving stereoscopic displays and for accommodation control," *Journal of Vision*, vol. 10, no. 8, pp. 22, 2010. 3

[21] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004. 3

[22] Hansung Kim and Adrian Hilton, "3D scene reconstruction from multiple spherical stereo pairs," *International journal of computer vision*, vol. 104, no. 1, pp. 94–116, 2013. 4