

# SILVR: A Synthetic Immersive Large-Volume Plenoptic Dataset

Martijn Courteaux

[Martijn.Courteaux@UGent.be](mailto:Martijn.Courteaux@UGent.be)  
Ghent University – imec  
Zwijnaarde, Oost-Vlaanderen  
Belgium

Julie Artois

[Julie.Artois@UGent.be](mailto:Julie.Artois@UGent.be)  
Ghent University – imec  
Zwijnaarde, Oost-Vlaanderen  
Belgium

Stijn De Pauw

[Stijn.DePauw1@gmail.com](mailto:Stijn.DePauw1@gmail.com)  
Ghent University – imec  
Zwijnaarde, Oost-Vlaanderen  
Belgium

Peter Lambert

[Peter.Lambert@UGent.be](mailto:Peter.Lambert@UGent.be)  
Ghent University – imec  
Zwijnaarde, Oost-Vlaanderen  
Belgium

Glenn Van Wallendael

[Glenn.VanWallendael@UGent.be](mailto:Glenn.VanWallendael@UGent.be)  
Ghent University – imec  
Zwijnaarde, Oost-Vlaanderen  
Belgium

## ABSTRACT

In six-degrees-of-freedom light-field (LF) experiences, the viewer's freedom is limited by the extent to which the plenoptic function was sampled. Existing LF datasets represent only small portions of the plenoptic function, such that they either cover a small volume, or they have limited field of view. Therefore, we propose a new LF image dataset “*SILVR*” that allows for six-degrees-of-freedom navigation in much larger volumes while maintaining full panoramic field of view. We rendered three different virtual scenes in various configurations, where the number of views ranges from 642 to 2226. One of these scenes (called *Zen Garden*) is a novel scene, and is made publicly available. We chose to position the virtual cameras closely together in large cuboid and spherical organisations ( $2.2\text{m}^3$  to  $48\text{m}^3$ ), equipped with  $180^\circ$  fish-eye lenses. Every view is rendered to a color image and depth map of  $2048\text{px} \times 2048\text{px}$ . Additionally, we present the software used to automate the multi-view rendering process, as well as a lens-reprojection tool that converts between images with panoramic or fish-eye projection to a standard rectilinear (*i.e.*, perspective) projection. Finally, we demonstrate how the proposed dataset and software can be used to evaluate LF coding/rendering techniques (in this case for training NeRFs with instant-nvgp). As such, we provide the first publicly-available LF dataset for large volumes of light with full panoramic field of view.

## CCS CONCEPTS

- Information systems → Multimedia databases; • Computing methodologies → Image processing; Image-based rendering.

## KEYWORDS

dataset, immersive, plenoptic, light field, 6DoF content

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*MMSys '22, June 14–17, 2022, Athlone, Ireland*

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9283-9/22/06...\$15.00

<https://doi.org/10.1145/3524273.3532890>

## ACM Reference Format:

Martijn Courteaux, Julie Artois, Stijn De Pauw, Peter Lambert, and Glenn Van Wallendael. 2022. SILVR: A Synthetic Immersive Large-Volume Plenoptic Dataset. In *13th ACM Multimedia Systems Conference (MMSys '22, June 14–17, 2022, Athlone, Ireland)*. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3524273.3532890>

## 1 INTRODUCTION AND MOTIVATION

In today's multimedia landscape, photos and videos are omnipresent and a lot of research and development activities are targeting more immersive ways to consume content. Using technologies such as Virtual Reality (VR), one can navigate with six-degrees-of-freedom (6DoF) through a virtual environment. This has led to a surge in research to methodologies for capturing and representing real-world environments, such that these can be be streamed and/or displayed in real time with 6DoF. Many such methodologies and systems start from images or videos of this environment as their source of data. The combination of many individual photographic captures of an environment, with accurate knowledge of the used lens and camera position for every image, is called a light field (LF) capture.

Examples that demonstrate the potential of using LF captures over traditional single-camera captures include Google's “Welcome To Light Fields” [17] and DeepView [4] and Meta's “Manifold in Action” [18]. These allow for high-quality, real-time rendering in VR, but the freedom of motion of the viewer is confined by a small sphere determined by the used capture setup. Enabling research for immersive 6DoF experiences with less confined freedom of movement relies on the availability of datasets that capture the environment in an immersive way over larger volumes. Unfortunately, to the best of our knowledge, no currently available dataset has the combination of properties required for large-volume 6DoF immersive navigation. For this reason, we propose a novel dataset “*SILVR*” (*Synthetic Immersive Large-Volume Ray* dataset) with the following features:

- (i) Three photo-realistic synthetic scenes, rendered with ground-truth depth maps. The scenes contain several elements that are known to pose challenges for many LF technologies, such as specularities, volumetrics, reflections, and translucency.
- (ii) High-resolution ( $2048\text{px} \times 2048\text{px}$ ) fish-eye ( $180^\circ$  field of view) cameras are evenly distributed ( $\pm 0.1\text{m}$  apart) across

the surface of a cuboid or sphere. This leads to a dense sampling of the plenoptic function, and thus of the light rays entering the cuboid or sphere.

- (iii) The cuboids and spheres are large ( $2.2\text{m}^3$  to  $48\text{m}^3$ ), with cameras looking outward. This allows for 6DoF in a large volume.

The dataset consists of static images, ground-truth depth maps, and configuration files containing details about the intrinsics and extrinsics of each camera. In addition to this, we make two tools openly available: the addon that was developed to render the images with Blender, and a reprojection tool that converts between panoramic or ultra-wide and rectilinear images. The dataset and tools can be found on <https://idlmedia.github.io/large-lightfields-dataset>, together with more in-depth information, examples and instructions on how to reproduce all results.

## 2 RELATED DATASETS

This section discusses some existing relevant LF datasets and assesses their applicability to large-volume immersive 6DoF experiences. Aspects such as the presence of depth maps, as well as the presence of optical elements such as specularities, volumetrics (e.g., fog, water, fire, smoke), and translucency are considered. Table 1 summarizes the datasets described below. Most datasets fall in one of three categories: *inside-out light fields*, *outside-in light fields*, *multiple panoramic 360° captures*. Broxton et al. introduced the term *interpolation volume* (IV) [6] (or *ray intersection volume* in [5]) as the volume behind the cameras that consists of the points for which incoming light rays can be obtained by interpolating data captured from the cameras [5, 6]. We extend this concept and introduce *immersive interpolation volume* (IIV), which is the interpolation volume limited to the points for which light rays in *any* direction can be obtained by interpolation. We are interested in datasets with a large IIV, as these enable immersive 6DoF experiences.

The first category of LF datasets uses cameras in an *inside-out* fashion to capture the environment. Many datasets in this category use camera setups organized on a plane [1, 8, 10, 12, 21] or curved surface. A notable example is the camera rig by Broxton et al., where 46 video cameras are mounted on an acrylic dome, which provides an IV which is around 80cm in width [5]. While some datasets have a considerable IV, they all have an empty IIV, as the cameras were all pointing towards the same region of the environment. Pozo et al. built a 16-camera rig with cameras positioned on a sphere pointing outward, yielding a spherical IIV with a diameter of approximately 1m, which they use to record immersive video [18]. Broxton et al. released their *Welcome to Light Fields* software demo on SteamVR along with several LF captures with a spherical IIV with a diameter of approximately 60cm [4, 17]. However, they use a customized closed-source VP9 video codec to compress the LF data [17], which makes it difficult to use it for research. Additionally, LFs captured with plenoptic cameras (e.g., Lytro Illum) like the Kalantari dataset [11] and *The Stanford Lytro Light Field Archive* [13] also belong in this category, but because they are not immersive (i.e., no 360° light information is available) and have a tiny IV (only a few centimeters wide), they are not included in Table 1.

The next category of LF datasets employs *outside-in* capturing for object in an environment, without explicitly capturing the environment itself. The cameras are positioned around the object and point inward to capture this object. As such, these LF captures are not considered immersive in the context of this paper. While the IV is typically large as it reaches entirely around the object, the IIV is empty. Examples of such datasets include [2, 15, 20, 22] which we will not further discuss but are included in Table 1 instead.

The last category of LF datasets uses multiple 360°-panoramic captures from an environment. These 360° photographs are typically taken on the same elevation from the ground and relatively far apart ( $\pm 2\text{m}$ ), such as the work by Chang et al. [7] and Armeni et al. [3] in which the *Matterport* camera is used. The Matterport camera is put on a rotating tripod and produces 18 images to cover the full horizontal 360° field of view. However, the poles (above and below the camera) are missing from the field of view, and the resulting images are of rather low resolution ( $1280\text{px} \times 1024\text{px}$ ), which likely results in noticeable aliasing in modern VR headsets. The FTV360 dataset by Maugey et al. captures videos from 40 tripod-mounted 360° cameras simultaneously [14]. The distance between neighboring cameras is between 1m and 3m and camera positions are estimated for all recorded sequences. The fact that the different viewpoints are far apart makes such datasets unsuitable for LF interpolation techniques, as these typically require denser sampled data. We reflect this shortcoming in Table 1 by assigning an empty IV ( $\emptyset$ ) for such datasets, although the size of the area they span with the camera positions is included between parenthesis for comparison. Techniques that do use such sparse data typically rely on accurate depth estimations, and use intermediate representations like meshes or point clouds. While these techniques deliver immersive 6DoF experiences, they rarely use LF concepts at their core. Additionally, when capturing an environment sparsely like this, challenging visual elements like specularities, volumetrics, and refraction effects are not adequately captured in the dataset.

We conclude that only the dataset from Pozo et al. [18] is truly immersive. However, while the IIV is 1m in diameter (which is among the largest of real-world camera setups), it consists of only 16 input views. This low number of views makes it difficult to adequately capture specularities and other difficult elements, as explained for the multiple-360° category. Overall, there are no public datasets capturing an environment in an inside-out way with a dense camera setup.

## 3 DATASET GENERATION AND DESCRIPTION

To address the absence of dense immersive light field datasets, this paper proposes a new dataset with a large interpolation volume covered by many cameras in an immersive configuration. To this end, three virtual scenes were rendered with Blender's ray-tracing engine *Cycles*. Section 3.1 discusses the scenes content-wise, and Section 3.2 describes the selected camera configurations, and the rendering process.

Dataset	Environment				Data properties				Note
		Natural / Synthetic	Specularities	Volumetrics / Translucencies	Resolution	Interpolation Volume	Immersive (Omni-directional/360°)	Has depth	
Inside-out LFs	Stanford LF Archive [12]	N	✓		various ( $\pm 1\text{MP}$ )	tiny			289 ( $=17 \times 17$ ) (1)
	OrangeKitchen [1]	S	✓	✓	1920 $\times$ 1080	tiny	✓		25
	NokiaChess [1]	S	✓		2048 $\times$ 2048	small	✓	✓	10
	Technicolor LF Dataset [21]	N		✓	2048 $\times$ 1088	small			16 ( $=4 \times 4$ )
	Google Spaces [8]	N	✓	✓	2048 $\times$ 1229	small			16
	Pozo et al. 6DoF Video Camera [18]	N		✓	3160 $\times$ 2160	medium	✓		16 (2)
	Hu et al. 4DLFVD [10]	N	✓	✓	1920 $\times$ 1056	small			100
	Broxton et al. [5]	N	✓	✓	2560 $\times$ 1920	medium			46 (3)
Outside-in LFs	Meta CO3D [20]	N	✓	✓	various	-			various
	Google RealEstate10K [22]	N	✓	✓	various	-			various (4)
	DTU Multi View Stereo [2]	N	✓		1600 $\times$ 1200	-			49 or 64
	NeRF (llff) data [15]	N	✓	✓	4032 $\times$ 3024	-			20 to 62
Multi 360°	ClassroomVideo [1]	S		✓	4096 $\times$ 2048	$\emptyset$ (medium)	✓	✓	15
	Matterport3D [7]	N	✓	✓	1280 $\times$ 1024	$\emptyset$ (medium)	✓	✓	18n (5)
	Armeni et al. 2D-3D-semantic [3]	N	✓	✓	1280 $\times$ 1024	$\emptyset$ (medium)	✓	✓	18n (5)
	Maugey et al. FTV360 [14]	N		✓	3840 $\times$ 1920	$\emptyset$ (large)	✓		40
Ours	<b>Barbershop</b>	S	✓	✓	2048 $\times$ 2048	large	✓	✓	642 or 1400
	<b>Lone Monk</b>	S	✓		2048 $\times$ 2048	large	✓	✓	642 or 2226
	<b>Zen Garden</b>	S	✓	✓	2048 $\times$ 2048	large	✓	✓	642 or 1806

**Table 1: Overview of existing scenes and datasets related to LF or immersive media.** Notes: (1) Tarot Cards has interesting light refraction through a glass ball. (2) 16 Fisheye lenses on a sphere with diameter 1m, yielding an IIV of approximately  $0.52\text{m}^3$ . (3) 46 Fisheye lenses on a dome. (4) YouTube videos solved for camera intrinsics and extrinsics, meaning mostly 1-dimensional camera traces. (5) For one panoramic image, 18 pictures are taken by a rotating camera on a tripod. The dataset consists of 10800 panoramas for Matterport3D and 1413 for 2D-3D semantic.

### 3.1 Virtual Scenes

Virtual scenes created in Blender were chosen over real-world environments. This decision was made for several reasons. First, rendering software enables a flexible configuration of camera setups and artistic choices, while still being able to produce photo-realistic results. Second, software lenses are perfect, do not require calibration, and do not require the camera positions and orientations to be known. Third, rendering can be done on existing general-purpose computer hardware. Finally, rendering is relatively inexpensive<sup>1</sup>.

Two of the scenes we selected are openly-available demo files from the Blender website<sup>2</sup>: “Agent 327: Barbershop” (CC-BY, Blender Foundation [9]) and “Lone Monk” (CC-BY, by Carlo Bergonzini from Monorender). Both scenes have enough geometry and decorations to make for an interesting immersive experience. This is in contrast to artists that only focus on the area that is visible by their chosen camera viewpoint. The Barbershop scene features the interior of a barbershop with many small props, fine detail, and a mirror. Not a single patch in this room looks not decorated. Lone Monk features

a courtyard of a cloister with a water well in the center. Small adjustments were made to make the scene slightly more immersive, as one side of the courtyard had some issues with the roof (these issues were not visible from the original camera). Additionally, a *solidify* modifier was added to the roof tiles to give them some thickness.

The third scene “Zen Garden” was built from scratch by the authors, using CC-0 models and textures. We integrated elements which are known to be challenging for LF technologies. More specifically, the statue and floor stones have reflective materials causing complex specularities. Additionally, there is a small mirror next to the statue. There is also a fire, several candles, and lanterns with a translucent case, all of which produce ambiguous depths. There is fog visible in front of the distant mountains and, finally, there are lots of small details such as tree branches, grass, and bamboo.

The three rendered scenes are shown in Fig. 1.

### 3.2 Camera Setup and Image Generation

In order to maximize the IIV for a given camera setup, (equidistant) fish-eye lenses with a 180° field of view are used. We chose to provide two camera-positioning setups per scene: one cuboid and one spherical, as demonstrated in Fig. 2. The spherical setup is an elegant way to capture lots of light rays from many directions entering

<sup>1</sup>With a total rendering time of around 12 days on a Linux system with two NVIDIA Quadro P5000 GPUs, consuming an average 500W (based on the on-device display), total energy usage is estimated to be around 144kWh. Associated electricity costs are estimated to be around 46 euro.

<sup>2</sup><https://www.blender.org/download/demo-files/>



**Figure 1: Overview of the Barbershop, Lone Monk, and Zen Garden scenes.**

the volume. The cuboid option is added to allow for cameras in six planar setups, which can yield even larger IIVs, as cuboids usually fit easier in closed areas, compared to a sphere. The specifics of these setups can be found in Table 2. In the table, it can be seen that the IIV is indeed significantly larger than any of the discussed datasets in Section 2, and that the number of views is high to maintain dense spatial sampling. Additionally, we rendered two extra evaluation configurations: a spherical setup with fewer cameras using perspective lenses, and another setup with 8 equirectangular panoramic renders as seen from the corners of a cube. These additional renders are provided to serve as ground truth data to evaluate view synthesis algorithms, as they do not coincide with any of the views in the other configurations. However, the fact that the viewpoints in the base configurations are so numerous and densely positioned enables researchers to subdivide them into input (*e.g.*, training) and evaluation sets if necessary. If a uniform division is desirable, we suggest to use the views with odd sequence numbers for evaluation.

In order to configure these camera setups easily and render all of the respective views, a novel Blender addon was developed. This addon is available online at <https://github.com/IDLabMedia/blender-lightfield-addon>. All resulting image files have a resolution of 2048px × 2048px with a circular projection and are stored losslessly in the OpenEXR file format, using 16-bit floating points for all color channels and the depth. Depth is measured in meters from the camera projection plane (meaning Z-depth, and not Euclidean distance). For each setup, all camera intrinsics and extrinsics are recorded in a JSON file.

## 4 REPROJECTING TO DIFFERENT LENS TYPES

As this dataset provides rendered views with fish-eye projections, these views are not directly suited to for applications requiring perspective projection views. To address this, we have developed an additional software tool that can convert images between different projections, assuming perfect lenses. This tool reads the EXR files, reprojects them according to the specified desired lens type and lens parameters, and writes the result back to EXR or PNG format. It should be noted that only EXR output format will also reproject the depth map, since PNG has no support for accurately storing floating point data required for depth values. Additionally, when using this tool, one should consider the artifacts of the reprojection process, such as aliasing and upsampling. Aliasing can be desirable in certain situations, but can be prevented by using multi-sampled reprojection (which can be configured through the command line argument `--samples`). Perspective projections stretch out the edges of the field of view enormously when using a large field of view. This can cause upsampling (*i.e.*, significantly more pixels cover a patch of the field of view in the perspective-projected image than were available in the fish-eye source data). We therefore suggest to either work with the fish-eye source data directly whenever possible, or consider using smaller field of views (which significantly shrinks the IIV). Alternatively, to avoid upsampling with large field of views, the result can be downsampled by using the command line argument `--scale` with an argument smaller than 1, which we recommend to combine with multi sampling to avoid aliasing in the center.

## 5 EXAMPLE USE CASE: NERF

In this Section, we describe the process of preparing the dataset for use in an example application, in this case NVLabs' Instant Neural Graphics Primitives (instant-nvg) implementation for training Neural Radiance Fields (NerF) [16]. While not fundamental to NeRFs themselves, instant-nvg expects images to use a perspective projection. As such, in the following example command, we convert the spherical camera setup of Lone Monk to a rectilinear lens (*i.e.*, lens with perspective projection) with focal length of 18mm, recorded on an image sensor of size 36mm × 36mm, stored as PNG files:

```
./reproject --parallel 4 --rectilinear 18,36 --scale 0.125 \
--png --exposure -1 --reinhard 5 \
--input-dir lone_monk/LFSphere_e220cm_d400cm/exr \
--input-cfg lone_monk/LFSphere_e220cm_d400cm/lightfield.json \
--output-dir lone_monk_perspective \
--output-cfg lone_monk_perspective/lightfield.json
```



**Figure 2: The placement of the cuboid and sphere camera setups in Zen Garden. The small black arrows indicate the direction each camera is facing. For the cuboid, all cameras belonging to the same face have the same rotation.**

			Number of views in total	Number of views along (X, Y, Z)	Dimensions	Mean distance between cameras
Barbershop	cuboid	1400	(10, 30, 10)	1m × 3m × 1m	11cm	
	sphere	642	-	1.45m diameter	9cm	
Lone Monk	cuboid	2226	(21, 21, 16)	4m × 4m × 3m	20cm	
	sphere	642	-	4.0m diameter	31cm	
Zen Garden	cuboid	1806	(21, 21, 11)	2m × 2m × 1m	10cm	
	sphere	642	-	1.7m diameter	13cm	

**Table 2: A summary of the camera setups used in the proposed dataset. For the dimensions of the cuboids, the third value indicates the vertical axis. The mean distance is calculated to each camera’s nearest (non-overlapping) neighbour.**

In the command, the lens type and specifics are extracted from the `lightfield.json` input configuration file. Note that the dimensions of the image are reduced to 1/8th (*i.e.*, 256px × 256px) and no anti-aliasing measures are taken (*i.e.*, there is no `--samples 8` flag). This exploits the fact that instant-npg can train the NeRF by only generating rays from the pixel centers and that we have a lot of different views available. Combining many aliased views still yields good results as the aliasing per view is high but the aliasing in the light field in its entirety is relatively low. Additionally, the EXR files store HDR content, and the Lone Monk scene is too bright to convert to PNG (which does not have HDR support) without adjusting exposure. As such, the exposure is reduced by one stop (`--exposure -1`) and Reinhard tone mapping with maximum brightness 5 (`--reinhard 5`) is applied [19].

Next, the configuration format for instant-npg is different and uses a different coordinate system. In the following example command, the generated configuration file from the previous step (that describes the reprojected dataset) is converted to the format of instant-npg:

```
python3 generate_NERF_transforms.py \
--scene lone_monk \
--dataset-config lone_monk_perspective/lightfield.json \
--output-transforms lone_monk_perspective/transforms.json
```

The used Python script is also available on our GitHub repository. Note that the ‘`--scene lone_monk`’ flag is used to use reasonable default values for the scaling and positioning of the scene within the NeRF bounding volume. After executing the two provided example commands, the generated ‘`lone_monk_perspective`’ folder is ready to be opened by instant-npg<sup>3</sup>.

## 6 CONCLUSION

In this paper, we presented a new light field dataset designed for six-degrees-of-freedom navigation with full panoramic vision in a large volume. Due to the large volume, objects can be visible, then occluded, and then visible again when moving linearly through the volume, as the viewpoint can fully traverse past an occluder, which is less common in previous datasets. Additionally, the digital scenes (with all used camera setups) and software tools we developed are made public. More specifically, three scenes (Blender’s “Agent 327 Barbershop”, “Lone Monk”, and our own novel “Zen Garden”) were rendered using a novel Blender plugin, using both a cuboid-

<sup>3</sup>A few animated results can be found at <https://idlabmedia.github.io/large-lightfields-dataset>.

and sphere-shaped camera setup. This resulted in the presented dataset of multi-view images and associated depth maps. Lastly, we presented software that converts images from fish-eye projection to rectilinear projection and vice versa, to broaden applicability of the dataset.

Although the camera setups used in this paper are large compared to the related work, the Barbershop, Lone Monk and Zen Garden scenes allow for even larger and unconventionally shaped setups. Zen Garden contains dynamic elements, such as the water and fire, so capturing videos instead of static images is possible. The dataset stems from virtual environments, which allows researchers to test their approaches with ground truth depth information available (when not ambiguous due to translucency). Future work includes extending the presented tools and software pipeline towards light field videos, which entails handling even more data and finding or creating interesting animated scenes.

## ACKNOWLEDGMENTS

This work was funded in part by the Research Foundation – Flanders (FWO) under Grant 1SA7919N, in part by IDLab (Ghent University – imec), in part by Flanders Innovation & Entrepreneurship (VLAIO), and in part by the European Union.

## REFERENCES

- [1] ISO/IEC JTC 1/SC 29/WG 11. 2020. Common Test Conditions for Immersive Video [N19214].
- [2] Henrik Aanæs, Rasmus Ramsbøl Jensen, George Vogiatzis, Engin Tola, and Anders Bjørholm Dahl. 2016. Large-Scale Data for Multiple-View Stereopsis. *Int. J. Comput. Vision* 120, 2 (nov 2016), 153–168. <https://doi.org/10.1007/s11263-016-0902-9>
- [3] Iro Armeni, Sasha Sax, Amir R Zamir, and Silvio Savarese. 2017. Joint 2d-3d-semantic data for indoor scene understanding. *CoRR* abs/1702.01105 (2017). arXiv:1702.01105 <http://arxiv.org/abs/1702.01105>
- [4] Michael Broxton, Jay Busch, Jason Dourgarian, Matthew DuVall, Daniel Erickson, Dan Evangelatos, John Flynn, Peter Hedman, Ryan Overbeck, Matt Whalen, and Paul Debevec. 2020. DeepView Immersive Light Field Video. In *ACM SIGGRAPH 2020 Immersive Pavilion* (Virtual Event, USA) (*SIGGRAPH ’20*). Association for Computing Machinery, New York, NY, USA, Article 15, 2 pages. <https://doi.org/10.1145/3388536.3407878>
- [5] Michael Broxton, Jay Busch, Jason Dourgarian, Matthew DuVall, Daniel Erickson, Dan Evangelatos, John Flynn, Ryan Overbeck, Matt Whalen, and Paul Debevec. 2019. A Low Cost Multi-Camera Array for Panoramic Light Field Video Capture. In *SIGGRAPH Asia 2019 Posters* (Brisbane, QLD, Australia) (*SA ’19*). Association for Computing Machinery, New York, NY, USA, Article 25, 2 pages. <https://doi.org/10.1145/3355056.3364593>
- [6] Michael Broxton, John Flynn, Ryan Overbeck, Daniel Erickson, Peter Hedman, Matthew DuVall, Jason Dourgarian, Jay Busch, Matt Whalen, and Paul Debevec. 2020. Immersive Light Field Video with a Layered Mesh Representation. *ACM Trans. Graph.* 39, 4, Article 86 (jul 2020), 15 pages. <https://doi.org/10.1145/3386569.3392485>

- [7] Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. 2017. Matterport3D: Learning from RGB-D Data in Indoor Environments. *CoRR* abs/1709.06158 (2017). arXiv:1709.06158 <http://arxiv.org/abs/1709.06158>
- [8] John Flynn, Michael Broxton, Paul E. Debevec, Matthew DuVall, Graham Fyffe, Ryan S. Overbeck, Noah Snavely, and Richard Tucker. 2019. DeepView: View Synthesis with Learned Gradient Descent. *CoRR* abs/1906.07316 (2019). arXiv:1906.07316 <http://arxiv.org/abs/1906.07316>
- [9] Blender Foundation. 2022. *Agent 327: Barbershop*. Blender Studio. <https://studio.blender.org/>
- [10] Xinjue Hu, Chenchen Wang, Yuxuan Pan, Yunning Liu, Yumei Wang, Yu Liu, Lin Zhang, and Shervin Shirnhammad. 2021. *4DLFVD: A 4D Light Field Video Dataset*. Association for Computing Machinery, New York, NY, USA, 287–292. <https://doi.org/10.1145/3458305.3478450>
- [11] Nima Khademi Kalantari, Ting-Chun Wang, and Ravi Ramamoorthi. 2016. Learning-Based View Synthesis for Light Field Cameras. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2016)* 35, 6 (2016).
- [12] Computer Graphics Laboratory. 2008. *The (New) Stanford Light Field Archive*. Stanford University. <http://lightfield.stanford.edu/lfs.html>
- [13] Computer Graphics Laboratory. 2008. *Stanford Lytro Light Field Archive*. Stanford University. <http://lightfields.stanford.edu/LF2016.html>
- [14] Thomas Maugey, Laurent Guillo, and Cedric Le Cam. 2019. FTV360: A Multiview 360° Video Dataset with Calibration Parameters. In *Proceedings of the 10th ACM Multimedia Systems Conference* (Amherst, Massachusetts) (MMSys '19). Association for Computing Machinery, New York, NY, USA, 291–295. <https://doi.org/10.1145/3304109.3325815>
- [15] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2021. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. *Commun. ACM* 65, 1 (dec 2021), 99–106. <https://doi.org/10.1145/3503250>
- [16] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. [arXiv:2201.05989](https://arxiv.org/abs/2201.05989) (Jan. 2022).
- [17] Ryan S. Overbeck, Daniel Erickson, Daniel Evangelakos, Matt Pharr, and Paul Debevec. 2018. A System for Acquiring, Processing, and Rendering Panoramic Light Field Stills for Virtual Reality. *ACM Trans. Graph.* 37, 6, Article 197 (dec 2018), 15 pages. <https://doi.org/10.1145/3272127.3275031>
- [18] Albert Parra Pozo, Michael Toksvig, Terry Filiba Schrager, Joyce Hsu, Uday Mathur, Alexander Sorkine-Hornung, Rick Szeliski, and Brian Cabral. 2019. An Integrated 6DoF Video Camera and System Design. *ACM Trans. Graph.* 38, 6, Article 216 (nov 2019), 16 pages. <https://doi.org/10.1145/3355089.3356555>
- [19] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. 2002. Photographic Tone Reproduction for Digital Images. *ACM Trans. Graph.* 21, 3 (jul 2002), 267–276. <https://doi.org/10.1145/566654.566575>
- [20] Jeremy Reizenstein, Roman Shapovalov, Philipp Henzler, Luca Sbdordone, Patrick Labatut, and David Novotny. 2021. Common Objects in 3D: Large-Scale Learning and Evaluation of Real-Life 3D Category Reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 10901–10911.
- [21] Neus Sabater, Guillaume Boisson, Benoit Vandame, Paul Kerbirou, Frederic Babon, Matthieu Hog, Remy Gendrot, Tristan Langlois, Olivier Bureller, Arno Schubert, and Valerie Allié. 2017. Dataset and Pipeline for Multi-view Light-Field Video. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 1743–1753. <https://doi.org/10.1109/CVPRW.2017.221>
- [22] Richard Tucker and Noah Snavely. 2018. *RealEstate10K*. Google. <https://github.com/realestate10k/>