CLASSROOM: synthetic high dynamic range light field dataset

Mary Guindy[a,b], Vamsi K. Adhikarla[b], Peter A. Kara[c,d], Tibor Balogh[a], and Aniko Simon[e]

[a]Holografika, Budapest, Hungary
[b]Pazmany Peter Catholic University, Budapest, Hungary
[c]Budapest University of Technology and Economics, Budapest, Hungary
[d]Kingston University, London, UK
[e]Sigma Technology, Budapest, Hungary

ABSTRACT

Light field images provide tremendous amounts of visual information regarding the represented scenes, as they describe the light traversing in all directions for all the points of 3D space. Due to the recent technological advancements of light field visualization and its increasing relevance in research, the need for light field image datasets has risen significantly. Among the applications for which light field datasets are considered, high dynamic range light field image reconstruction has gained notable attention in the past years. When capturing a scene, either a single camera with a 2D microlense array or a 2D array of cameras is used to produce narrow- and wide-baseline light field images, respectively. Additionally, the turn-table methodology may be used as well for narrow-baseline light fields. While the majority of these methods enables the creation of plausible and reliable light field image datasets, such baseline-specific setups can be extremely expensive and may require immense computing resources for proper calibration. Furthermore, the resulting light field is commonly limited with regard to angular resolution. A suitable alternative to produce a light field dataset is to do it synthetically by rendering light field images, which may easily overcome the aforementioned issues. In this paper, we discuss our work on creating the "CLASSROOM" light field image dataset, depicting a classroom scene. The content is rendered in horizontal-only parallax and full parallax as well. The scene contains a high variety of light distribution, particularly involving under-exposed and over-exposed regions, which are essential to HDR image applications.

Keywords: Light field imaging, 3D rendering, HDR, dataset

1. INTRODUCTION

Light field (LF) images provide tremendous amounts of visual information regarding the represented scenes, as they describe the light traversing in all directions for all the points of 3D space. At the time of writing this paper, light field displays (LFDs) are already available, but they have not emerged on the consumer market yet; to reach that milestone, much scientific work is still left to be done. Generally speaking, an LFD acts as a window to the 3D world described by the corresponding light rays.[1] Although this technology is considered to be a breakthrough in 3D visualization by many – as the presented contents are perceivable without the need of additional viewing devices – yet it has immense requirements on multiple fronts simultaneously. Depending on the field of view (FOV), contents must be captured from lots of different angular perspectives in order to provide an accurate representation. This corresponds to a 4D function representation of LFs in the case of free occluder space, where the spatial and angular information are both recorded representing the different perspectives of the scene from multiple viewing points. For an LF scene to be rendered, multiple images are required, since a single image corresponds to a slice in the 4D LF representation.[2] Thus, in order to create an LF image dataset, multiple images are needed for a single LF scene. The creation of an LF dataset depends on many factors:

Further author information: (Send correspondence to Mary Guindy)
Mary Guindy: E-mail: m.guindy@holografika.com / guindy.mary.mohsen.messak@ppke.hu
Vamsi K. Adhikarla: E-mail: adhikarla.vamsi.kiran@itk.ppke.hu
Peter A. Kara: E-mail: kara@hit.bme.hu / p.kara@kingston.ac.uk
Tibor Balogh: E-mail: t.balogh@holografika.com
Aniko Simon: E-mail: aniko.simon@sigmatechnology.se

- System baseline: This considers the FOV of the LF system. Whereas a narrow-baseline system usually has an FOV between 10° and 15°, a wide-baseline system has an FOV that is equal to or greater than 30°.
- Parallax: Regarding the different LFDs, they can be categorized based on their parallax. Horizontal-only parallax (HOP) displays show the scene from different angular perspectives along the horizontal axis. Vertical-only parallax (VOP) displays are possible on a technical level, yet less relevant on a practical level, since the human eyes are horizontally separated. Full-parallax (FP) displays support both horizontal and vertical parallax, and accordingly, they are significantly more challenging to design and implement than HOP and VOP displays.
- Distance between the LFD and the observer line / rectangle: For HOP systems, we consider the possible positions where the observer can view the screen (i.e., the observer line), whereas for FP systems, the same is considered but in both directions (horizontally and vertically), hence, the observer rectangle.
- Arrangement of optical elements: This is concerned with wide-baseline systems. For narrow-baseline devices, usually the micro-lens array within the sensor is arranged horizontally or in a 2D-manner without difference in orientation. On the other hand, capturing wide-baseline LFs requires cameras to be arranged in a 1D or a 2D array with the possibility of aligning the cameras in an arc according to the optical system. In that case, the orientation of the cameras are different from one another, resulting in images with slight rotations.

With the recent technological advancements of LF visualization and its increasing relevance in research, the need for LF image datasets is quite considerable; however, since a single content is represented by multiple LF images describing it from various perspectives, such datasets are extremely huge compared to conventional image datasets. Moreover, capturing scenes for LFDs is also challenging and expensive due to the nature of the potential options. For narrow-baseline LF contents, capturing is achieved either via a single camera with a 2D array of sensors or the turntable methodology, during which a conventional camera captures a scene or an object that rotates around its vertical axis. For wide-baseline contents, a 2D array of cameras is used to capture the content. While the majority of these methods enables the creation of plausible and reliable LF image datasets, such baseline-specific setups can be rather expensive and may require immense computing resources for proper calibration. Furthermore, the resulting LF is commonly limited with regard to angular resolution. A suitable alternative to produce an LF dataset is to create it synthetically by rendering LF images, which may easily overcome the aforementioned issues. Among the applications for which LF datasets are considered, High Dynamic Range (HDR) LF image reconstruction has gained notable attention over the past years. In addition to the creation of HDR contents for LFDs, the need for Low Dynamic Range (LDR) to HDR LF conversion must also be eventually performed for the legacy LF contents. This could be done via Convolutional Neural Networks (CNNs), hence the need for an HDR LF dataset that can be used for training and testing. Some datasets already exist for LF images, yet very few of them consist of HDR LF images, which additionally increases the need for such datasets.

In this paper, we discuss our work on creating the "CLASSROOM" HDR LF image dataset, depicting a classroom scene. The content is synthetically created and it is rendered with the help of a virtual camera. Analogous to LF capture by using a 2D array of cameras, rendering for the scene is done multiple times with a slight camera movement from one position to another. The virtual camera relocates horizontally to capture the different perspectives of the scene, and the changes along the vertical axis are neglected since the dataset is particularly designed for the current HOP displays. However, in order to support future FP displays, the dataset is extended by a complete FP version. In practice, the horizontal and vertical switches of camera positions are precise enough to produce the amount of images required to fully capture the scene without suffering from the perceivable artefacts while observing the visualized content on an LFD. In the rendering process, we consider linear arrangements of 2D camera arrays; in other words, no camera rotations are done due to the sparsity of arc LFDs. Regarding the contents of the scene, a high variety of light distribution is considered, particularly involving under-exposed and over-exposed regions, which are essential to HDR image applications. The resolution values of the dataset are fundamentally set by the capabilities of the state-of-the-art, high-end LFDs, yet at the same time, they are limited by the training and testing times of the CNNs. The resulting dataset is evaluated by experts of the field; however, utilizing the dataset in practical applications and performing a series of subjective tests are out of the scope of the paper.

The remainder of the paper is structured as follows: Section 2 discusses the previously created datasets for the LF and HDR LF systems, as well as, the different HDR formats. Section 3 discusses the reasons behind choosing the classroom scenario for the dataset and the synthetic nature of rendering, and also introduces the setup used for its creation. Section 4 shows the results for the created datasets, and finally, the work is concluded in Section 5, with possible scientific directions for future work.

2. RELATED WORK

According to Metzler *et al.*,[3] many techniques can be used to reconstruct HDR images from LDR images, including reverse tone mapping methods, computational photography methods and CNNs. Among those techniques, CNNs have proven to provide the best results with the ability to further improve. Similarly, CNNs can be used for LDR-to-HDR LF image reconstruction. In our previous paper,[4] we have tested different CNNs on HDR LF images to visualize the results in order to proceed with the next steps for better HDR LF image reconstruction. However, one of the main challenges was the lack of HDR LF datasets. Since the outputs of deep learning depend on both the deep complex structures of the networks and the large training datasets,[5] acquiring more HDR LF datasets will further improve the research on HDR LF image reconstruction.

2.1 Light field datasets

Due to its increasing importance in various applications, LF imaging has become essential in several research fields. As mentioned in Section 1, LF captures more information about the scene compared to conventional imaging, since spatial and angular information are both recorded. Hence, LF datasets are larger than the conventional image datasets, as a single scene is represented by multiple images.[6] Multiple attempts were done in creating LF datasets, among which is the multiview HOP dataset with a single object in the scene,[7] the SMART dataset with 15 LF images,[8] the dense LF dataset with 14 scenes using 5 synthetic images,[9] the VALID dataset,[10] and a 10-scene dataset with 5 degrees of freedom.[11]

2.2 HDR image formats and encoding

As a means of storing HDR images, HDR image formats have emerged, recording wider color gamuts compared to RGB images. These formats take into consideration several aspects, including file size, total dynamic range and the size of the smallest step between the consecutive values. Among the different HDR image formats are HDR, TIFF and EXR. The HDR format (.hdr and .pic) was first introduced in 1989, covering more than 76 orders of magnitude, with files as large as uncompressed 24-bit RGB images, since the used run-length encoding achieves 25% compression rate. Compared to HDR encoding, the TIFF float format takes almost three times the storage space, since floating numbers are not well compressed. On the other hand, it is best suited for writing and reading float-point frame buffers. Since users always favor compressed files for easier usage and storage, the LogLuv encoding was introduced for a more compact TIFF representation.[12] Later in 2002, EXR (Extended Range format) was introduced as an open source C++ library used for reading and writing EXR images. In EXR, both the 16-bit and 32-bit floating point numbers are used for storing pixel data. The EXR format supports mipmapping, tiling, as well as lossless compression. For compression, either ZIP deflate library or Industrial Light and Magics (ILM) are used, with the latter being more compression efficient, resulting in a 60% compression. Moreover, EXR supports random channels such as user-defined ones, alpha, depth, etc.[12,13]

2.3 Towards HDR LF datasets

Combining both LF and HDR technologies is rather powerful, where 3D content is visualized with an added sense of realism, close to the HVS. However, the creation of an HDR LF dataset requires tremendous amounts of storage due to the aforementioned reasons in Sections 2.1 and 2.2.

To the best knowledge of the authors, only a single dataset considers HDR LF imaging. The "Teddy" HDR LF dataset[14] was captured by means of a DSLR camera. The capturing setup was mounted on a moving stage, allowing sub-mm precision in a range of 4 m horizontally and 0.5 m vertically. For HDR image creation, the exposure bracketing method was used,[15] where multiple LDR images with different exposures are captured and then merged together to reconstruct a single HDR image.

3. THE CLASSROOM DATASET

3.1 Reasons for creating the dataset

Although the “Teddy” dataset provides insightful uses in the HDR LF reconstruction field, it is still a single dataset, which cannot be solely used to evaluate the performance of a reconstruction CNN. Moreover, it is hardware customized, thus considering only a single use case. On the other hand, the CLASSROOM dataset is synthetically created, hence, the ability to change different parameters and adapting it to different conditions, while creating more datasets. Also, increasing the scene complexity is possible by adding more objects or upgrading the scene to have more complex materials. This can be useful in the progressive learning curve of the HDR LF reconstruction field. In addition to the aforementioned reasons, a synthetic dataset is not custom-designed to a certain type of baseline or parallax, as different alterations can be made to render multiple datasets for different baseline and parallax settings.

The reason behind choosing the classroom scene is due to its ability to provide HDR images. The concept of HDR relies on having a big dynamic range of colors in the scene. Considering the classroom scene, this is possible since there are areas where light penetrates the classroom windows, creating over-exposed regions, whereas on the other hand, some regions in the classroom (e.g., cupboards and bookshelves) are under-exposed, hence the high dynamic color range in the produced images.

3.2 MAYA setup

In order to create the CLASSROOM dataset, we used MAYA (version of 2022). For rendering the modeled classroom, the Arnold renderer was used. This renderer is an advanced Monte Carlo ray tracing renderer, which is both memory-efficient and scalable. Multiple features are integrated in the Arnold renderer, including – but not limited to – subsurface scatter, hair and fur, motion blur, volumes, instances, subdivision and displacement, OSL support, light path expressions, adaptive sampling, toon shader and – most importantly – denoising, which was used as a post-processing step in the dataset to eliminate the noise resulting from the Monte Carlo algorithm. Due to its efficiency and plausible results, the Arnold renderer is integrated in many softwares, such as MAYA, Houdini, Cinema 4D, 3Ds Max and Katana.[16] In addition to the aforementioned capabilities, the Arnold renderer allows the usage of Image-based lighting (IBL), firstly introduced by Debevec[17] in 2008, which allows synthetic objects to be rendered in real-world scenes. In other words, illumination and lighting can be used from real-world scenes by means of HDR images (HDRI) to illuminate the modeled synthetic scenes, adding realism to the output content. In order to use IBL for realistically illuminating the classroom scene, an HDRI was imported from the “polyhaven” website[*] (previously named HDRI haven). With the rising importance of HDR imaging, the newer versions of MAYA support HDR formats. In the CLASSROOM dataset, we used OpenEXR 32-bit floating point images.

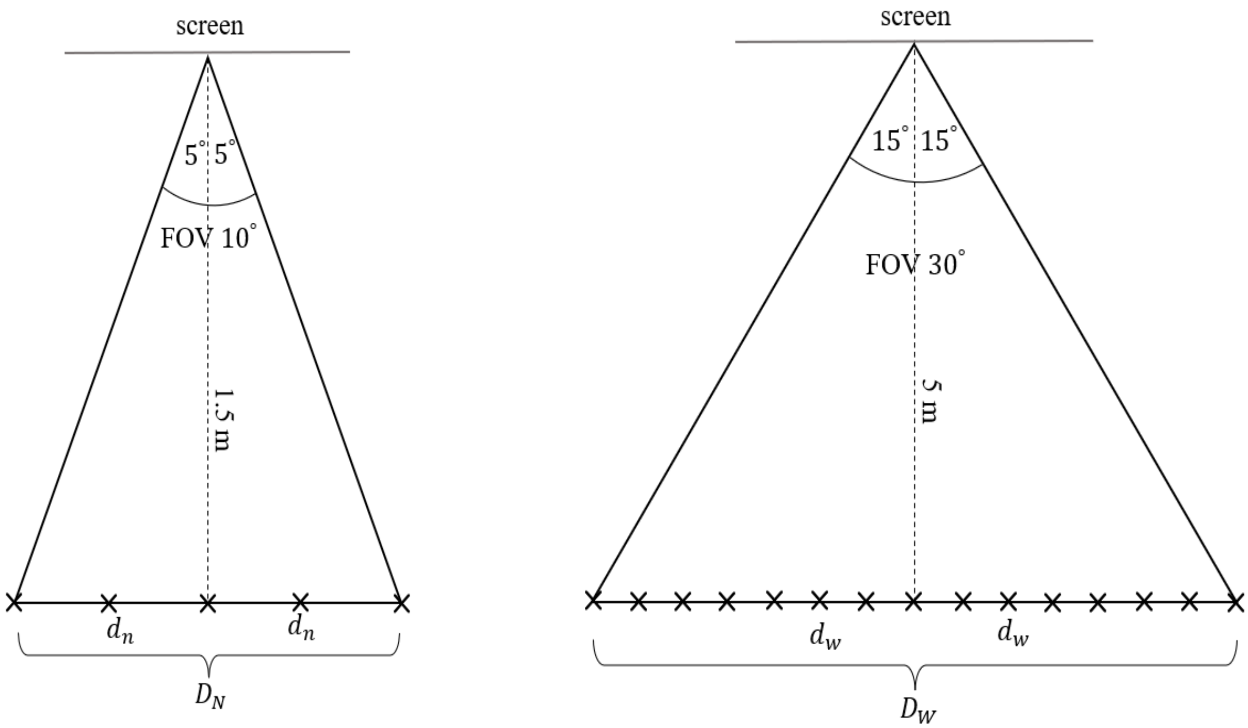
3.3 Distance calculation

For the creation of the CLASSROOM dataset, we consider both baselines (i.e., narrow- and wide-baseline systems), the distance between the LFD and the observer line or rectangle and both parallax cases (i.e., HOP and FP). In order to understand the reason behind the chosen distances between the consecutive images in the dataset, we consider the following two cases.

The first case considers a narrow-baseline LF system with an FOV of 10° and a distance of 1.5 m between the observer line or rectangle and the LFD screen. This is illustrated in Figure 1a, depicting the top view of the LF system setup. The distance d_n can be calculated as $1.5 * tan(5°) = 0.1312m \approx 13cm$, with a total distance $D_N = 26cm$. In the narrow-baseline FP dataset, we consider 5 images in each direction, hence, the distance between any two consecutive images in the horizontal or vertical directions is $26/4 = 6.5cm$.

For the wide-baseline LF system, we consider an FOV of 30°, with a distance of 5 m between the observer line or rectangle and the LFD screen. Accordingly from Figure 1b, the distance d_w can be calculated as $5 * tan(15°) = 1.339m \approx 133cm$, with a total distance $D_W = 266cm$. For the wide-baseline HOP dataset, a total of 15 images were rendered, therefore, the distance between each two consecutive images is $266/14 = 19cm$.

[*] https://polyhaven.com/



(a) Top view for narrow-baseline LF.

(b) Top view for wide-baseline LF.

Figure 1: Distance calculation in narrow- and wide-baseline LF systems.

4. RENDERED RESULTS

The CLASSROOM dataset consists of three subsets: (i) narrow-baseline FP, (ii) narrow-baseline HOP and (iii) wide-baseline HOP. The images are rendered using an Intel(R) Core(TM) i7-5820K CPU with 6 cores. For all datasets, we consider an image size of 960×540. The reason for the chosen size is to avoid having small-sized images (i.e., loss of details) and large-sized images (i.e., too much time and complexity when applying HDR LF reconstruction techniques). The creation of the components of the scene (e.g., chairs) followed a public online tutorial on Autodesk Maya[†].

4.1 Narrow-baseline FP dataset

Starting off with the narrow-baseline FP dataset, we created 25 images arranged in a 5×5 2D array. The distance between each two consecutive images is 6.5 cm in both the horizontal and vertical directions, covering a total distance of 26 cm spanned in the 10° FOV of the considered narrow-baseline system. The camera used for creating narrow-baseline datasets had a focal length of 35 mm. Figure 2 illustrates the distances between the rendered images with respect to the FOV.

The final rendered images for the dataset are illustrated in Figure 3, with the image EXR file size ranging between 27.1 MB and 30 MB and a total size of 713 MB per dataset. The time taken to render a single image ranged between 6:13 min and 6:56 min with an average of 6:31 min per image.

[†]Hassaan Owaisi: Classroom interior modeling in maya
https://www.youtube.com/watch?v=lRrLqR_5eBM

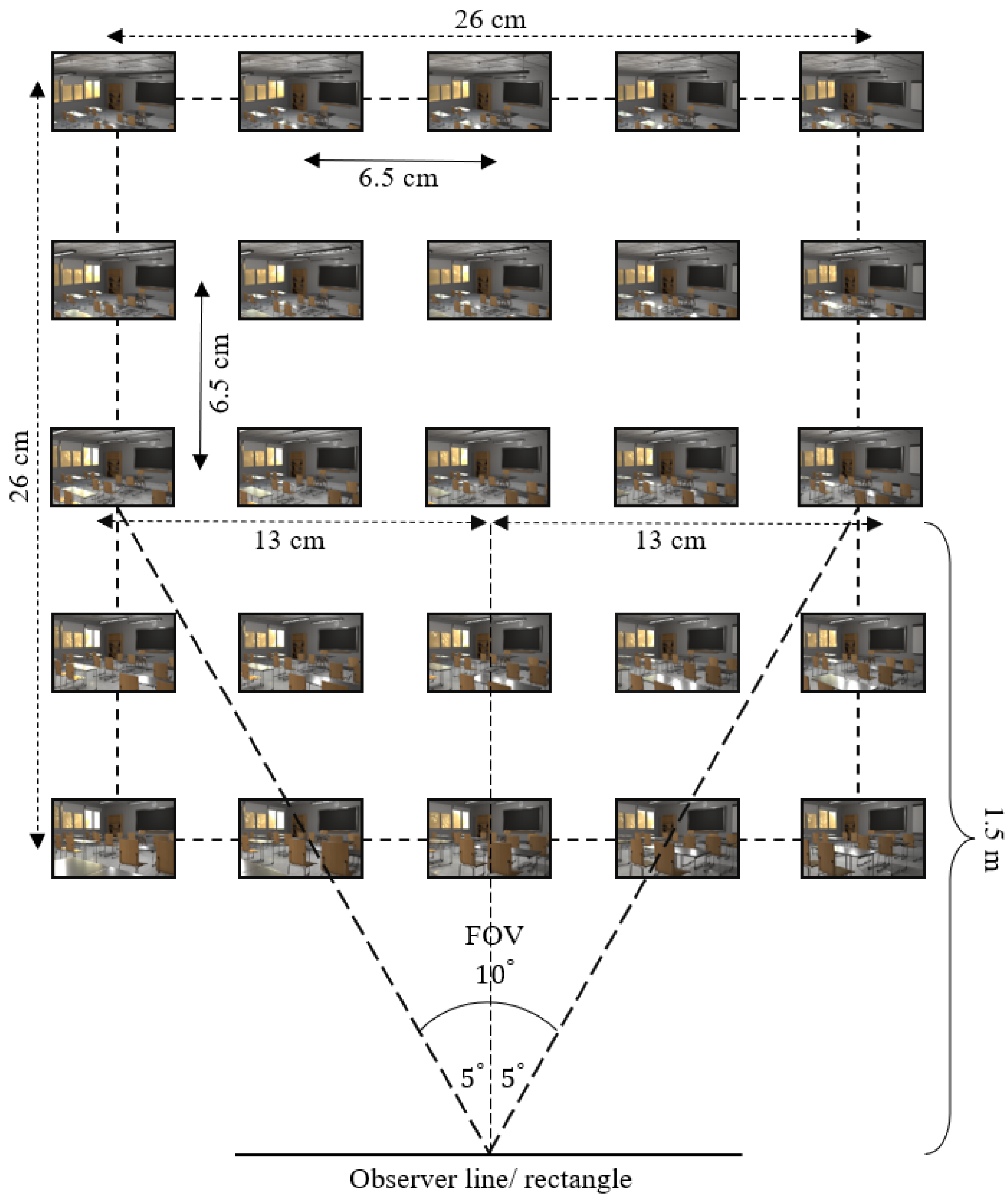


Figure 2: Narrow-baseline FP dataset setup.

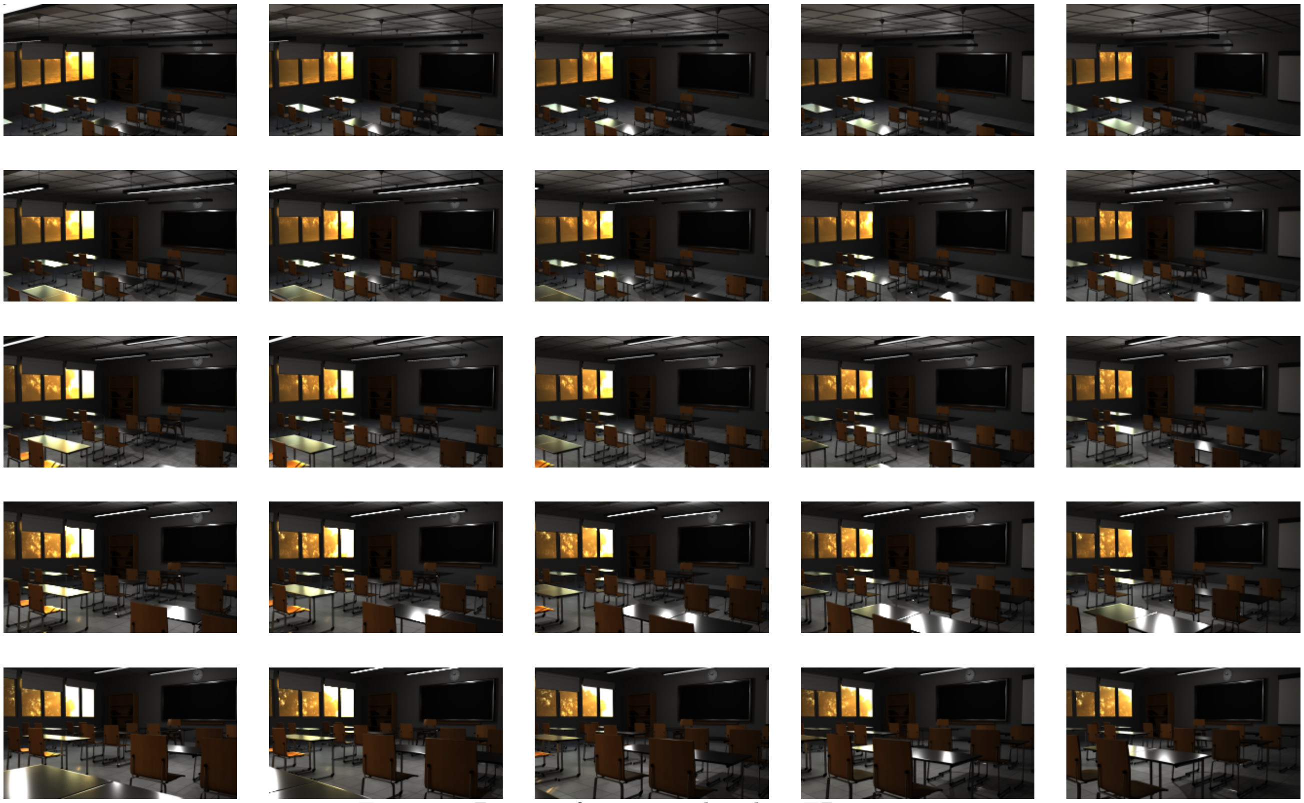


Figure 3: Dataset for narrow-baseline FP systems.

4.2 Narrow-baseline HOP dataset

The narrow-baseline HOP dataset is considered to be a subset of the narrow-baseline FP dataset, since the HOP considers horizontal directions only. Accordingly, given the narrow-baseline FP dataset, 5 datasets can be created for the HOP system, as illustrated in Figure 4.

4.3 Wide-baseline HOP dataset

In this dataset, we consider wide-baseline HOP systems, rendering a total of 15 images for the dataset arranged in a 1D horizontal array. Figure 5 depicts the relation between the rendered images and the FOV of the wide-baseline systems, where the distance between any two consecutive images is 19 cm, thus, covering a total distance of 266 cm spanned by the wide-baseline system with an FOV of 30°. For rendering, a camera with a focal length of 20 mm was used to allow for wider motions in the scene.

Figure 6 shows the rendered images constituting the wide-baseline HOP dataset, where images are arranged from right to left and top to bottom. The image EXR file size ranges between 19.3 MB and 27.5 MB with a total size of 370 MB for the dataset. The time taken to render an image ranged between 5:18 min and 6:54 min with an average of 6:14 min per image.

5. CONCLUSION AND FUTURE WORK

With the importance of HDR and LF imaging in different applications, combining them leads to powerful results. Hence, the importance of reconstructing HDR LF images from the legacy LDR LF images is evident. This, however, requires different HDR LF datasets upon which the different reconstruction CNNs can be applied and tested. In this paper, we presented our work on creating a dataset for HDR LF applications. The dataset incorporates three datasets, targeted for different LF systems: (i) narrow-baseline FP, (ii) narrow-baseline HOP and (iii) wide-baseline HOP.

Figure 4: Narrow-baseline HOP datasets from the FP dataset.

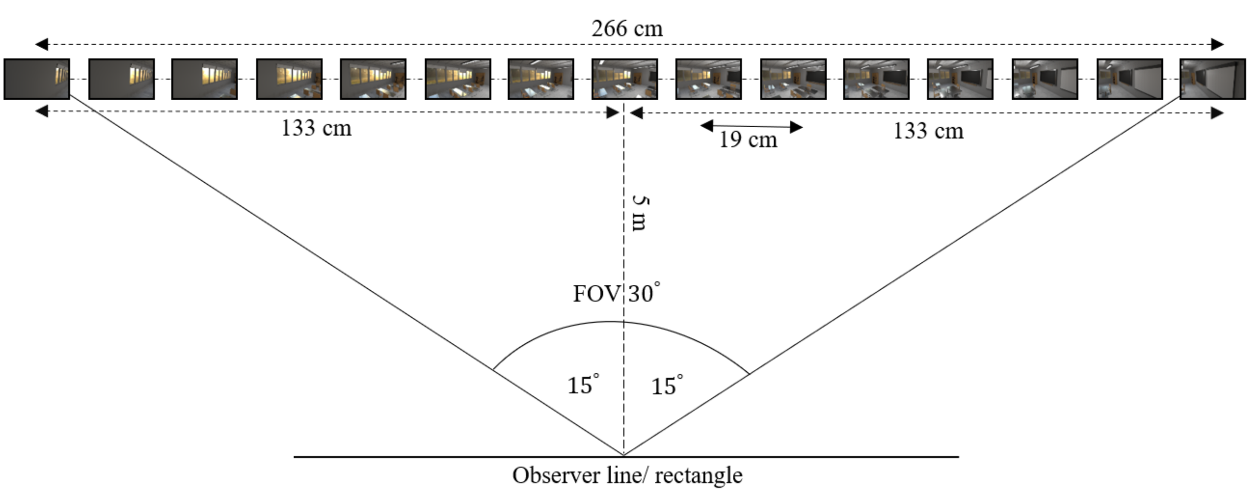


Figure 5: Wide baseline dataset setup

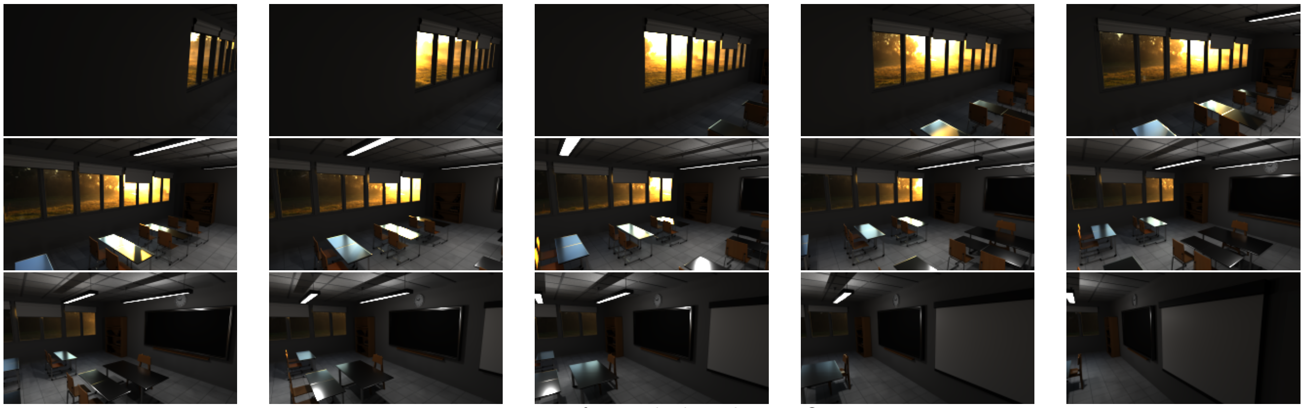


Figure 6: Dataset for wide-baseline HOP systems

The classroom scene rendered in the output images includes the bare minimum details. This is due to the fact that the HDR LF image research has just recently started, so it would be better to start with fewer details and lower complexity. With the progressive HDR LF reconstruction learning curve, increasing scene complexity is possible by adding more objects or using more complex materials. In addition to increasing the complexity, a dataset for arc systems can be created by using a camera with the aim (in MAYA) to render images with different orientations.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813170. Also received funding by 2018-2.1.3-EUREKA-2018-00007 and KFI 16-1-2017-0015, NRDI Fund, Hungary; and by project no. BME-NVA-02, implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021 funding scheme.

REFERENCES

[1] Balram, N. and Tošić, I., "Light-field imaging and display systems," *Information Display* **32**(4), 6–13 (2016).

[2] Levoy, M. and Hanrahan, P., "Light field rendering," in [*Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*], 31–42 (1996).

[3] Metzler, C. A., Ikoma, H., Peng, Y., and Wetzstein, G., "Deep optics for single-shot high-dynamic-range imaging," in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*], 1375–1385 (2020).

[4] Guindy, M., Kiran, A. V., Kara, P. A., Balogh, T., and Simon, A., "Performance evaluation of HDR image reconstruction techniques on light field images," in [*2021 International Conference on 3D Immersion (IC3D)*], 1–7, IEEE (2021).

[5] Wang, W., Zhang, M., Chen, G., Jagadish, H., Ooi, B. C., and Tan, K.-L., "Database meets deep learning: Challenges and opportunities," *ACM SIGMOD Record* **45**(2), 17–22 (2016).

[6] Ellahi, W., Vigier, T., and Le Callet, P., "Analysis of public light field datasets for visual quality assessment and new challenges," in [*European Light Field Imaging Workshop*], (2019).

[7] Tamboli, R. R., Appina, B., Channappayya, S., and Jana, S., "Super-multiview content with high angular resolution: 3D quality assessment on horizontal-parallax lightfield display," *Signal Processing: Image Communication* **47**, 42–55 (2016).

[8] Paudyal, P., Battisti, F., Sjöström, M., Olsson, R., and Carli, M., "Towards the perceptual quality evaluation of compressed light field images," *IEEE Transactions on Broadcasting* **63**(3), 507–522 (2017).

[9] Adhikarla, V. K., Vinkler, M., Sumin, D., Mantiuk, R. K., Myszkowski, K., Seidel, H.-P., and Didyk, P., "Towards a quality metric for dense light fields," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 58–67 (2017).

[10] Viola, I. and Ebrahimi, T., "Valid: Visual quality assessment for light field images dataset," in [*2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*], 1–3, IEEE (2018).

[11] Shi, L., Zhao, S., Zhou, W., and Chen, Z., "Perceptual evaluation of light field image," in [*2018 25th IEEE International Conference on Image Processing (ICIP)*], 41–45, IEEE (2018).

[12] Reinhard, E., Heidrich, W., Debevec, P., Pattanaik, S., Ward, G., and Myszkowski, K., [*High dynamic range imaging: acquisition, display, and image-based lighting*], Morgan Kaufmann (2010).

[13] Kainz, F., Bogart, R., and Hess, D., "The openexr image file format," *ACM SIGGRAPH Technical Sketches* (2003).

[14] Gul, M. S. K., Wolf, T., Bätz, M., Ziegler, M., and Keinert, J., "A high-resolution high dynamic range light-field dataset with an application to view synthesis and tone-mapping," in [*2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*], 1–6, IEEE (2020).

[15] Debevec, P. E. and Malik, J., "Recovering high dynamic range radiance maps from photographs," in [*ACM SIGGRAPH 2008 classes*], 1–10 (2008).

[16] "Autodesk Arnold, tech. rep., https://damassets.autodesk.net/content/dam/autodesk/www/pdfs/arnold-features.pdf."

[17] Debevec, P., "Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography," in [*ACM SIGGRAPH 2008 classes*], 1–10 (2008).