CIVIT DATASETS: HORIZONTAL-PARALLAX-ONLY DENSELY-SAMPLED LIGHT-FIELDS

Sergio Moreschini, Filipe Gama, Robert Bregovic and Atanas Gotchev

Faculty of Information Technology and Communication Sciences
Tampere University, Finland

ABSTRACT

In this paper, we present datasets consisting of six horizontal-parallax-only densely-sampled light fields (DSLFs). Five of the presented light fields (LFs) have been captured using a camera attached to an in-house built motorized linear positioning system (LPS). The sixth LF has been generated in Blender by replicating the LPS' camera setup. All LFs have undergone pre- and post-processing steps in order to mitigate capture-related distortions. The generated LFs are organized, based on their characteristics, into three categories: synthetic, Lambertian and non-Lambertian LFs. All of them comply with the DSLF requirement, i.e. the maximum disparity between adjacent images is less than one pixel. The datasets are provided to serve as ground truth LFs for developing various applications, such as LF view interpolation, super resolution, and compression.

Index Terms — light field, dataset, non-Lambertian, LPS

1. INTRODUCTION

Light information within an observable space is completely described by a seven-dimensional function referred to as Plenoptic function [1]. These seven dimensions are composed of spatial positions, direction, time and wavelength of light rays. In practice, one can drop time in favor of time sequences and sample the color information of a ray through the RGB channels. Furthermore, the ray propagation is assumed only in front of the observer and not behind (i.e. half space). With these simplifications and in an occlusion-free environment, every light ray can be described by the intersection of two parallel planes. Such discretized 4D approximation of the Plenoptic function is referred to as the light field (LF) [2]. A specific case of the LF is the so called densely sample LF (DSLF) [3]. In a DSLF, the maximum displacement of any object point is no more than one pixel between two adjacent angular view images. DSLF is operational for synthesizing any desired ray by simple linear interpolation and therefore it is a preferred LF representation for a number of applications such as view rendering, super resolution, compression, transmission and visualization. Having ground truth (GT) DSLF is vital for developing these applications, which motivates the attempts to render or capture DSLFs with diverse scene characteristics.

LF capture systems are usually categorized into two categories: microlenses-based devices and gantry-based systems. The first category is composed of cameras that have a microlens array placed between the main lens and the image sensor of the camera [4] and are based on the so-called Plenoptic 1.0 [5] and Plenoptic 2.0 [6] principles. A Plenoptic 1.0 system has the microlens placed in the focal plane of the main lens and therefore each microlens captures one spatial location and multiple angles (i.e. the corresponding images have higher angular and lower spatial resolution). A Plenoptic 2.0 system has the microlenses focused on the image plane of the main lens, which generates images with higher spatial and lower angular resolution on the sensor [7].

The second category is composed of systems where single or multiple cameras are mounted on motorized or static rigs with the aim to capture scenes with wide baseline [8]. Multiple cameras on a single rig are used to capture dynamic scenes and a single camera on a motorized rig can capture a static DSLF.

For the aforementioned two categories, different datasets have been generated and made available to the research community. In the first category, some of the datasets captured with a Plenoptic 1.0 (Lytro) camera have been presented in [9] and [10]. Similarly, using a Plenoptic 2.0 (Raytrix) camera, a dataset of both real and synthetic (rendered) images has been presented in [11]. In order to provide fair comparison between the two available Plenoptic configurations in [12], a dataset captured with both Lytro and Raytrix camera has been presented.

In the second category the most well known dataset is the one by Stanford that was generated using two different gantry systems: a multi-camera array and a LF microscope [13]. For what concerns a single camera setup, usually the scene captured is of small dimensions for both resolution and baseline, as in [14], an exception being the dataset presented in [15] where the captured scenes have a significant change in perspective. A different kind of example is [16], where some of the LFs composing the dataset are synthetic scenes generated in Blender [17]. Synthetic scenes have different properties compared with real-life LF such as the complete absence of noise in the capturing stage or the possibility of generating purely Lambertian scenes.

In this work we present a DSLF dataset which belongs to the second category. The dataset contains both real and synthetic scenes. The wide baseline ensures significant changes in perspective. The dense sampling makes this dataset very appealing for analysis and in particular for LF reconstruction. In particular, it allows a straightforward visualization on LF displays, a more precise analysis of LF in frequency domain and testing depth estimation algorithms.

The paper is structured as follows: In Section 2, we introduce the used acquisition system composed of a camera and a gantry system. In Section 3, we present calibration and post-processing steps applied to the acquired data. In Section 4, we present the different scenes captured. Conclusions are presented in Section 5.

2. ACQUISITION SYSTEM

The acquisition system used to generate the dataset is composed of two main parts: A motorized linear positioning system (LPS) and a camera. The LPS is composed of an anti-backlash nut mounted on a hard alloy aluminium base powered by two stepper motors - one for horizontal and one for veritcal direction (Figure 1). Its precise linear movement and high accuracy in positioning ($\pm 20 \mu m$)

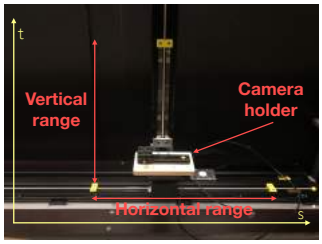
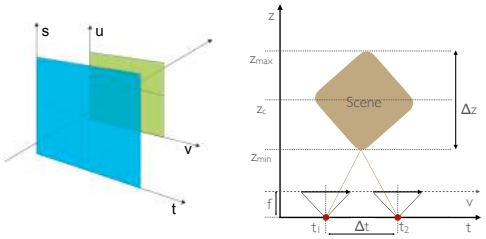


Figure 1: Motorized linear acquisition positioning system.



(a) (b)

Figure 2: LF formalization. (a) Two-plane parameterization. (b) Capture setup.

allowed us to capture the DSLF with a sub-pixel resolution precision. More information about the system can be found in [18].

The camera used for capture was an Optronis CP70-12-C-188 with Nikon Fx 35 mm f/1.8 lens. The camera can capture images up to a resolution of 4.080 x 3.072 pixels.

Following the LF formalization we define the camera plane as (s, t) and the image plane as (u, v) (Figure 2 (a)). Assuming that the camera moves over a single trajectory t, see figure Figure 2 (b), we can set s as a constant value achieving an imagery referred to as Horizontal Parallax Only (HPO) LF. By making use of the LPS, the camera moves in space along a line with a fixed step size Δt and for each step captures a picture of a stationary scene. For n images, it spans a width $T = \Delta t \cdot (n - 1)$.

3. DATASET ACQUISITION

The images captured using the setup described in Section 2 are affected by multiple undesired effects which need to be removed. First, various image corrections have to be applied, e.g. correction for lens distortion, color correction. Second, one needs to take into account small inaccuracies related to the difficulty of pointing the camera exactly orthogonal with respect to the moving axis. Third, the step size (distance camera moves between two consecutive images) has to be correctly selected with respect to the scene. A diagram depicting the different steps necessary for the creation of the dataset is shown in Figure 4. All those required steps can be performed by proper calibration of the capture system, as described in the next section.

3.1. System Calibration

The purpose of calibration is to correct color inaccuracies, lens distortion and camera orientation.

Color correction is performed by making use of a color checker. The *X-rite ColorChecker* [19] has been captured and, after performing demosaicing, color values have been extracted. From the difference between the extracted values and those indicated as correct, a color correction matrix has been generated.

The lens distortion has been computed by making use of a checkerboard. Multiple images of the planar pattern at different orientation have been captured. The distortion parameters have been computed based on the corners of the squares in the checkerboard [20].

Finally, from the captured multiple pictures of the checkerboard, projection of 3D points have been measured and, to estimate the camera parameters, the inverse problem has been solved through a nonlinear optimization algorithm [21].

3.2. Capture

Capture is designed in terms of scene composition and the camera step selection ensuring the DSLF requirement. Following the LF formalization introduced in Section 2, when a camera captures two subsequent pictures at a distance Δt at the camera plane, a point in the scene is represented at two different positions in the two captured images. The difference in coordinates of such a point in the images is defined as *disparity* d in pixels, and can be computed as (Figure 2 (b)):

$$d = \frac{\Delta v}{Pv} = \frac{v_1 - v_2}{Pv} = \frac{f}{z}\frac{(t_1 - t_2)}{P_v} = \frac{f}{z} \cdot \frac{\Delta t}{P_v}, \quad (1)$$

where P_v the size of the pixel on the image plane. P_v is evaluated as $P_v = \frac{2f}{N_{px}} \tan(\frac{FoV}{2})$ where N_{px} and FoV are the resolution on the v axis and the field of view, rispectively. Establishing a DSLF condition means ensuring that the maximum disparity, between any consecutive two images in the LF, is less or equal to 1. Assuming a recentering performed in post-processing, that is, ensuring that objects in center at distance $z_c = (z_{max} + z_{min})/2$ have disparity zero, it is important to have a difference in the disparity between the closest (z_{min}) and farthest (z_{max}) point in the scene of less than 2 pixels, that is

$$\Delta d = d_{zmin} - d_{zmax} \leq 2. \quad (2)$$

To this goal, one must first select a distance z_{min} at which the scene with depth Δz is placed. This can be done by selecting the smallest distance at which the overall scene is still visible in all captured images. Second, one needs to compute the correct step size between the adjacent camera position on the camera plane Δt which will ensure that the DSLF condition given by the equation (2) is satisfied. This can be done by making use of the following equation:

$$\Delta t = \frac{(z_{min} + \Delta z) \cdot z_{min}}{f} Pv \cdot \Delta d \quad (3)$$

3.3. Post-processing

Corrections are applied to all images in the captured dataset. The correction parameters are estimated in the calibration step. At first the color transform computed in the color calibration step is applied by exploiting the color correction matrix. Second, the lens distortion is corrected by making use of the lens correction matrix. Third, the images are rectified based on the estimated camera parameters. Finally, the images are recentered and cropped to the desired resolution.

The process of recentering involves the use of the shearing properties in the Epipolar Plane Image (EPI) domain. Following the formalization in Section 2, and assuming an HPO LF $H(u, v, t)$, an LF slice $E(v, t) = H(u_0, v, t)$, for fixed $u = u_0$, is referred to as EPI [22]. Each EPI is composed of lines of different directions. Each line describes an object point trajectory related with

the camera move. An object which stays in the center of the image along all of the views in the LF, is represented in the EPI domain as a vertical stripe. Therefore, the process of recentering involves finding the amount of shearing to be applied to all EPIs composing the LF, such that the stripe describing the object in the center of the physical scene is exactly vertical.

A set of all processed images for a given scene is the DSLF dataset for that scene.

4. DATASET

Six different scenes have been captured using the previously described capturing setup. Each of these scenes can be grouped into 3 different categories: synthetic, Lambertian and non-Lambertian. The synthetic scene, depicted in Figure 3 (a), is called *Toys*. The Lambertian scenes are those referred to as *Seal and Balls* and *Dragon* presented respectively in Figure 3 (b) and (c). Finally, the non-Lambertain scenes are *Flowers*, *Castle* and *Holiday* shown in Figure 3 (d), (e) and (f).

Each of the following dataset is composed of 193 images having resolution of 720x1280 pixels in the 8-bit color depth representation. Central views of all dataset are shown in Figure 3.

1. **Toys**: This scene is a synthetic scene generated in Blender [17] and mimics the movement of the capture system described in Section 2. Main challenges of this scene, both from the reproduction and the analysis point of view, are due to a high level of occlusions present in the scene.
2. **Seal and Balls**: This scene contains four plushies. Three of these are spheres while the last one is a seal. The particularity of this scene is the small amount of reflections generated when the light is impinging on the surface of the objects. This makes the Seal and Balls dataset a good example of a real scene with (near) Lambertian properties.
3. **Dragon**: This scene is characterized by multiple LEGO characters. Although some reflection are present in the scene, they are constant for the whole scene reducing to minimum the effect of the non-Lambertianity. Therefore, as in the case of the Seal and Balls scene, this can be categorized as a Lambertian scene.
4. **Flowers**: This scene is composed of multiple highly reflective surfaces and semi-transparent petal flowers. Moreover, most of the elements are represented by using mostly one of the three channels in the RGB representation. This makes the three color channels very different from each other.
5. **Castle**: In this scene, a castle made of LEGO with a multitude of LEGO soldiers and objects such as stairs are present in the scene. The light when impinging on some of these surfaces produces a very high reflection which completely cancels the properties of the color of the object making the scene highly non-Lambertian.
6. **Holiday**: This scene contains three main objects: a bottle, a lantern and a Christmas wreath, with multiple occlusions presented in the scene. The Christmas wreath is composed of multiple highly reflecting surfaces which project the light that hits on the bottle. Furthermore, the lantern contains multiple light sources, some of them occluded by the boundaries of the lantern itself. Such characteristics make this a very challenging dataset for LF reconstruction.

The dataset was firstly introduced as part of the ICME 2018 Grand Challenge on Densely Sampled Light Field Reconstruction [23]. The scenes: *Seal and Balls*, *Castle* and *Holiday* were part of the development dataset (DD) provided to the proponents, while the remaining three were part of the evaluation dataset (ED) that has been used for the evaluation of the proposed algorithms.

Link to dataset: http://urn.fi/urn:nbn:fi:att:ed60be6d-9d15-4857-aa0d-a30acd16001e

5. CONCLUSIONS

In this paper we presented a dataset for HPO DSLFs. Most of the scenes have been captured using a motorized LPS and a camera, the exception being the scene *Toys* generated in Blender. Proper calibration and post-processing steps have been performed in order to generate DSLFs suitable for further experimentation. When preparing the dataset, the emphasis was put on making DSLF based on the geometrical distances related to the objects of interest in the scene. However, due to parts of the scene being non-Lambertian (e.g. light reflections in the scene referred as *Holiday*) and having objects in the images not belonging to the scene (e.g. background), the 1 pixel DSLF requirement might not have been satisfied for all points in the imagery. The datasets are intended for LF analysis and are particularly suitable for evaluating the quality of LF reconstruction algorithms.

6. ACKNOWLEDGMENT

The work in this paper was funded by the European Union's Horizon 2020 research and innovative program under the Marie Sklodowska-Curie grant agreement No. 676401, European Training Network on Full Parallax Imaging and supported by CIVIT (Center for Immersive Visual Technologies).

7. REFERENCES

[1] E. Adelson and J. Bergen, "The plenoptic function and the elements of early vision," *Computational Models of Visual Processing,* vol. 1, MIT Press, 1991.

[2] M. Levoy and P. Hanrahan. "Light field rendering," *Proc. 23rd Annu. Conf. Compu. Graphics Interactive Techn.,* 1996, pp. 31-42.

[3] Z. Lin and H.-Y. Shum, "A Geometric Analysis of Light Field Rendering," *Int'l J. of Computer Vision*, vol. 58, no. 2, pp. 121138, 2004.

[4] G. Lippmann, "Epreuves reversibles donnant la sensation du relief," *J. Phys. Theor. Appl.*, vol. 7, no. 1, pp. 821-825, 1908.

[5] R. Ng, M. Levoy, B. Mathieu, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, vol. 2, no. 11, pp. 1-11, 2005.

[6] C. Perwass, and L. Wietzke, "Single lens 3d camera with extended depth-of-field," *Human Vision and Electronic Imaging XVII* International Society for Optics and Photonics, 2012, vol. 8291, p. 829108.

[7] B. Goldluecke, O. Klehm, S. Wanner, and E. Eisemann, (2015). "Plenoptic Cameras." *Digital Representations of the Real World: How to Capture, Model, and Render Visual Reality*, pp. 65-78, 2015.

[8] Y. Xu, K. Maeno, H. Nagahara, and R. I. Taniguchi, "Mobile camera array calibration for light field acquisition." *arXiv preprint arXiv:1407.4206*, 2014.

[9] P. Paudyal, R. Olsson, M. Sjöström, F. Battisti, and M. Carli, "Smart: A light field image quality dataset," *Proc. of the 7th Int. Conf. on Mult. Sys.*, p. 49, 2016.

[10] M. Rerabekand and T. Ebrahimi, "New light field image dataset," *Proc. 8th Int. Conf. on Quality of Multimedia Experience (QoMEX)*, 2016, number EPFL-CONF-218363.

[11] L. Palmieri, R. Op Het Veld, and R. Koch, "The Plenoptic 2.0 Toolbox: Benchmarking of depth estimation methods for MLA-based focused plenoptic cameras," *Proc. 25th IEEE International Conference on Image Processing (ICIP)*, pp. 649-653, 2018.

[12] W. Ahmad, L. Palmieri, R. Koch, and M. Sjöström, "Matching light field datasets from plenoptic cameras 1.0 and 2.0," *Proc. 3DTV Conference*, June, 2018.

[13] The (New) Stanford Light Field Archive, "https://lightfield.stanford.edu/lfs.html".

[14] K. Honauer, O. Johannsen, D. Kondermann and B. Goldluecke, "A Dataset and Evaluation Methodology for Depth Estimation on 4D Light Fields," *Asian Conference on Computer Vision*, Springer, pp. 19-34, 2016.

[15] M. Ziegler, R. Op Het Veld, K. Keiner and F. Zilly, "Acquisition System for Dense Lightfield of Large Scenes," *Proc. 3DTV-Conference*, pp. 1-4, 2017.

[16] V. Kiran Adhikarla, M. Vinkler, D. Sumin, R. K. Mantiuk, K. Myszkowski, H. P. Seidel, and P. Didyk, "Towards a quality metric for dense light fields," *Proc. of the IEEE Conf. on Comp. Vis. and Pattern Recognition*, pp. 58-67, 2017.

[17] Blender Foundation, "Free and open source 3D animation suite," 3D rendering software, www.blender.org.

[18] S. Vagharshakyan, A. Durmush, O. Suominen, R. Bregovic, and A. Gotchev, "Accuracy evaluation of a linear positioning system for light field capture," *Proc. Intelligent Information and Database Systems: 7th Asian Conference,* p. 388-397, Mar. 2015.

[19] X-rite ColorChecker Classic "https://xritephoto.com/colorchecker-classic"

[20] Z. Zhang. "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol, 22, 2000.

[21] R. Hartley, and A. Zisserman. *Multiple view geometry in computer vision,* Cambridge university press, 2003.

[22] R. Bolles, H. Baker and D. Marimont. "Epipolar-plane image analysis: An approach to determining structure from motion," *Int. J. Comput. Vis.,* vol. 1, no. 1, pp. 7-55, Mar. 1987.

[23] ICME 2018 - IEEE International Conference on Multimedia and Expo, "http://www.icme2018.org/conf_challenges".

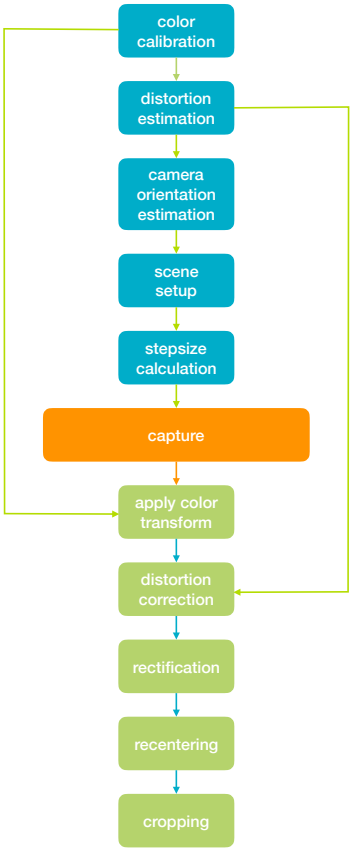


Figure 4: Diagram showing the different steps performed to create the dataset.

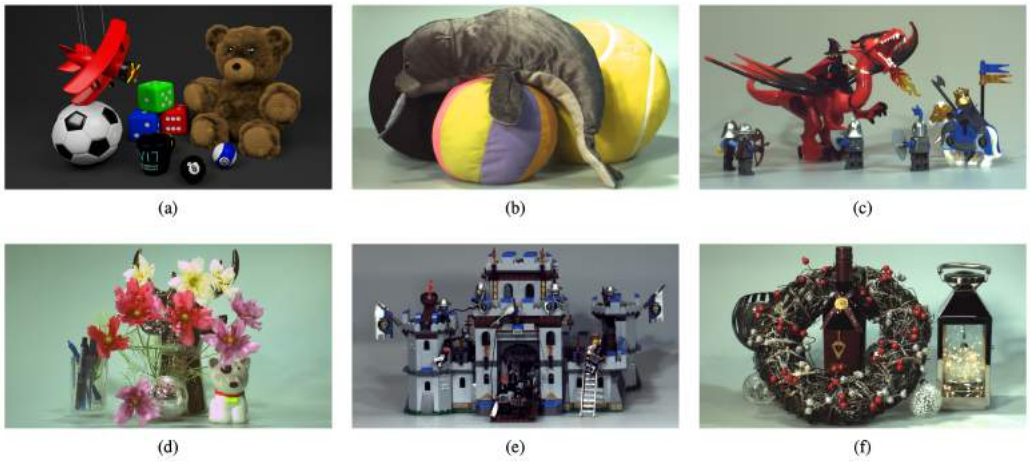


Figure 3: Dataset. (a) Toys. (b) Seal and Balls. (c) Dragon. (d) Flowers. (e) Castle. (f) Holiday.