Non-Planar Inside-Out Dense Light-Field Dataset and Reconstruction Pipeline

Faezeh Sadat Zakeri[1], Ahmed Durmush[2], Matthias Ziegler[1], Michel Bätz[1], and Joachim Keinert[1]

[1]Moving Picture Technologies, Fraunhofer IIS, Erlangen, Germany
[2]3D Media Group, Tampere University, Tampere, Finland

ABSTRACT

Light-field imaging provides full spatio-angular information of the real world by capturing the light rays in various directions. This allows image processing algorithms to result in immersive user experiences such as VR. To evaluate, and develop reconstruction algorithms, a precise and dense light-field dataset of the real world that can be used as ground truth is desirable. In this paper, a non-planar capture is done and a view rendering pipeline is implemented. The acquired dataset includes two scenes that are captured by an accurate industrial robot with an attached color camera such that the camera is looking outward. The arm moves on a cylindrical path for a field of view of 125 degrees with angular step size of 0.01 degrees. Both scenes and their corresponding geometric calibration parameters will be available with the publication of the paper. The images are pre-processed in different steps. The disparity between two adjacent views with resolution of $5168x3448$ is less than 1.6 pixels; the parallax between the foreground and the background objects is less than 0.6 pixels. Furthermore, the pre-processed data is used for a view rendering experiment to demonstrate an exemplary use case. In addition, the rendered results are evaluated visually and objectively.

Index Terms**—** Light-field; non-planar; dense dataset; inside-out capture; view rendering

1. INTRODUCTION

Light-field technology has opened many spaces for researchers to develop efficient light-field processing algorithms that provide realistic and immersive features which are not reachable by conventional photography.

In 1908, Lippmann [1] introduced the concept of integral imaging which led to today's light-field imaging that offers 3D information in an array of 2D images with different perspectives. To test, evaluate, and improve different algorithms such as depth estimation methods, depth imaged-based rendering techniques, quality assessment algorithms as well as developing algorithms for applications like VR, 360 stitching, 3D modeling, and many more, a precise and dense high resolution dataset is required where the camera positions are not located on a single plane. One main use case of such a dataset is environmental scanning that needs images to be captured in a non-planar fashion with high overlap in various camera poses. Note that environmental scanning requires an inside-out capture configuration where the optical axis of the camera is looking outward as the idea is to record the surroundings. Such data can be rendered using physical-based rendering pipelines for instance in Blender [2]. However, the simulation cannot completely resemble the properties of a real-world scene [3]. To the best of our knowledge, there is a lack of such real-world data.

Our captured dataset consists of two very dense, high resolution light-fields with sufficient change in perspective for non-planar camera positions. Both light-fields cover a field of view of around 125 degrees showing a large scene with a variety of different realistic objects. The first light-field contains difficult objects with properties such as transparencies or reflections for more advanced experiments. The second light-field, however, contains mostly less demanding objects and thus serves as a foundation for basic experiments. Due to this variety, our dataset is beneficial for a wide range of applications. Several images of the dataset are shown in Fig. 1.

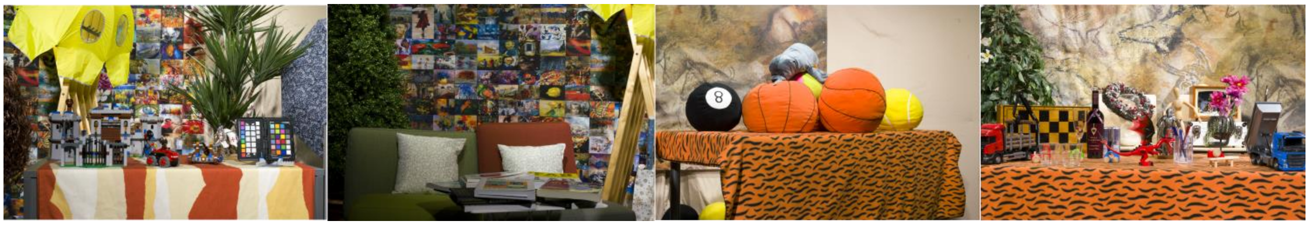


Figure 1: Examples of images in different positions of the both scenes

F. S. Zakeri, A. Durmush, M. Ziegler, M. Bätz and J. Keinert, "Non-Planar Inside-Out Dense Light-Field Dataset and Reconstruction Pipeline," 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 1059-1063.
doi: 10.1109/ICIP.2019.8803402, URL: https://ieeexplore.ieee.org/document/8803402

Currently, there is a variety of light-field datasets publicly available with different characteristics that are detailed in [4].
In most of them [5, 6] a 1D linear camera system is employed where the cameras configured in one direction only. Consequently, these datasets are not suitable for applications that need images of the scene in horizontal and vertical directions such as light-field displays.

There are also 2D light-field datasets that are captured using a 2D camera array where the cameras are uniformly spaced. [7, 8]. These camera arrays have the problem that they cannot capture very dense light-fields since the angular resolution is restricted by the number of cameras used and their physically minimum distance between cameras.

In addition to arrays, there are gantry systems that make capturing dense light-field possible. In [9], a system of two industrial cantilever axis and a camera that is mounted on the slider is employed. The camera can be positioned within the gantry in both horizontal and vertical direction by command control. These acquisition systems are designed for linear capturing, i.e., the orientation of the camera is perpendicular to the sampling plane. Besides, not all of the available datasets captured by such a setup provide images with enough density and high parallax in a high spatial resolution.

In this paper, we present our contribution by introducing a novel non-planar dataset and a view rendering pipeline. We highlight the specifications of our dataset. In addition, we illustrate the conceptual details of our rendering pipeline for the algorithm development.

The reminder of this paper is as following: in Section 2, the acquisition system plus the scene configuration are described. Afterwards, a brief explanation of our pre-processing is detailed in Section 3. In Section 4, the properties of the dataset are illustrated in detail. Our non-planar rendering pipeline and the applications of the dataset is discussed in Section 5. Finally, Section 6 concludes the paper.

2. LIGHT-FIELD AQCUISITION SYSTEM

2.1. Mechanical Setup

The industrial robotic arm *Kuka kr 16 L6-2* [10] was employed for capturing. As shown in Fig. 2, the camera is mounted on the sixth joint of the robotic arm such that the tripod mount point of the camera coincides with the center of rotation of the sixth joint.
To get enough parallax and change in perspective while moving the camera, the first joint is set as the center of rotation such that the length of the robotic arm is 1.5 meters relative to the first joint. This basically defines the radius of the circle that the camera moves on via rotating the arm. The robot permits a position repeatability of ±0.05 mm, a rotation repeatability of 0.005 degrees, and an assessed payload of 6 kg which allows mounting of an RGB camera with a lens. The robot has a wide range of rotation around its first joint that is detailed in [10]. This permits us to rotate the camera around the Z axis precisely and capture images in a non-planar manner, as shown in Fig. 3.

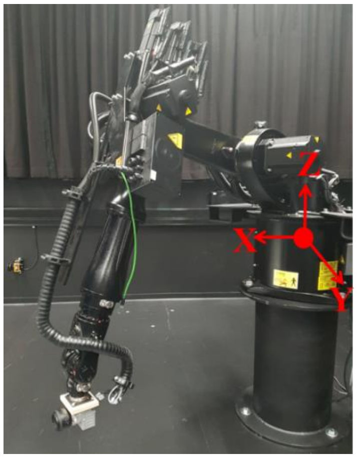


Figure 2: *Kuka kr* robotic arm is utilized for the light-field acquisition.

2.2. Scene Setup

Dense light-field capturing while having enough change in perspective at the same time is challenging. We achieved these two features by controlling four items: scene dimension, camera positioning steps, the distance of the scene relative to camera initial position, and the radius of the movement. Fig. 3 shows a model of our scene configuration. The distance of the closest and farthest object to the camera are 1.5 m and 3.5 m respectively. Thus, the total focus distance is around 2.5 m. The radius of the sampling path is 1.5 m. The camera moving path covers a 96 degree of rotation around Z axis. Considering the field of view of the lens, the total captured field of view is about 125 degrees.

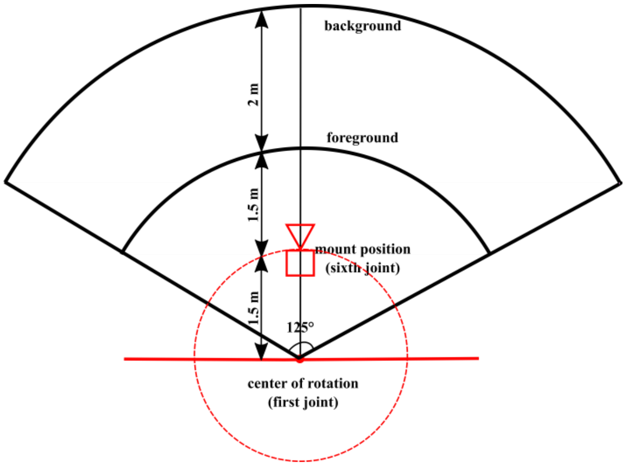


Figure 3: Illustration of the scene configuration. Relative distances and the path of the camera movement are shown.

3. PRE-PROCESSING CHAIN

We imported the raw captured image data into a pre-processing chain that includes six steps:

1. De-bayering,
2. Chromatic aberration correction,
3. Color correction,
4. Lens distortion correction,
5. Geometric calibration,
6. Compression.

Image_ID	qw	qx	qy	qz	tx	ty	tz	Camera_ID	Name
1	0.999519	-0.000199384	0.031008	0.000111161	6.20567	0.026925	-0.716325	1	0000_000.tif

Camera_ID	Model	Width	Height	Params
1	SIMPLE_RADIAL	5168	3448	7954.56, 2584, 1724, -0.00337946

Figure 4: An example of output files for images and cameras produced by Colmap as calibration data. Top pattern shows the content style of *images.txt*. Bottom pattern shows the content style of *cameras.txt*.

The steps 1-2 were done using Adobe Photoshop CS6. To do color correction, we first calibrate the camera's color profile by capturing an image of an X-rite color checker in the same lighting system as used for the whole experiment and then, profiled it using the ColorCheckerPassport software [11]. Subsequently, we imported the obtained camera's color profile into Photoshop CS6 as a preset and applied the color correction on all images. In order to remove lens distortion, we calibrated our camera using checkerboard images and undistorted all images via the MATLAB Calibration Toolbox.

Geometric calibration of the non-planar images was done using a structure-from-motion strategy for 3D reconstruction from an unordered image collection that is explained in [12] via the Colmap interface [13]. Colmap outputs two files in text format: *camera.txt* and *images.txt*. The structure of the *image.txt* file shown in the top of Fig. 4 is explained as following:

- Each line contains the parameters of a single view,
- The quaternion q holds the view orientation parameters,
- The camera position is encoded in vector t.

Similarly, the *cameras.txt* includes camera related fields as shown in the bottom of Fig. 4. The parameters in the vector params contains specific values for a certain camera model. The parameter interpretation of the output file produced by Colmap is as following: focal length, principal point \ [x and y and z].

The given calibration information with the datasets allows us to use the images for different purposes. For example, the geometric calibration data can be used to compute rectifying homographies given a pair of views.

Eventually images are compressed using HEVC (lossless mode). This reduces the size of first light-field from 2.4 TB to 0.9 TB and the second light-field from 1.4 TB to 0.5 TB.

4. LIGHT-FIELD DATASET DESCRIPTION

Our dataset includes two scenes in total. Fig. 1 shows several images of both scenes in different positions. The first scene: *Mannequin* consists of 9600x5 images. *Mannequin* scene is very complex and includes many Lambertian and non-Lambertian objects such as highly reflecting surfaces, shiny ones as well as transparencies and semi-transparencies. The scene also contains fine structures such as leaves or hair, homogeneous regions and periodic patterns, highly textured objects and those with high contrast, and shadows. The second scene: *Sofa*, consists of 9600x3 images. It is a simpler version of the *Mannequin* scene with less non-Lambertian and challenging objects. The scenes are captured in raw file format with a resolution of 5168x3438 pixels by a high quality mirrorless color camera (Sony alpha 7RII) equipped with a 35 mm Canon lens. Both scenes are configured such that there is 2 meters depth between the foreground and the background, Figure 2. The objects are placed in different depth layers and all are in focus. There are objects which intentionally placed such that they occlude other objects in both scenes to bring challenges for post processing.

5. APPLICATION

To prove our claim regarding the applications for this dataset, we have developed a depth-based non-planar view rendering pipeline. This pipeline considers the non-planar camera system where the cameras can have arbitrary positions and rotations. The implemented concept of this pipeline assumes a virtual camera partner for each camera as is shown in Fig. 5 (Cameras with white color). The virtual camera partner is defined the same as the actual camera. But, it is positioned away from the actual camera such that the baseline between the two cameras is one. Based on multi-view geometry detailed in [14], we create a projection matrix for each camera using the given geometric calibration information. Using the projection matrices, we apply stereo matching and estimate disparity maps per each view in our camera system.

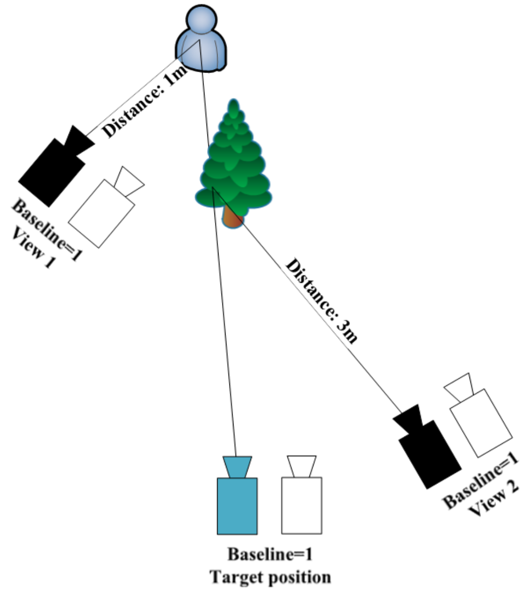
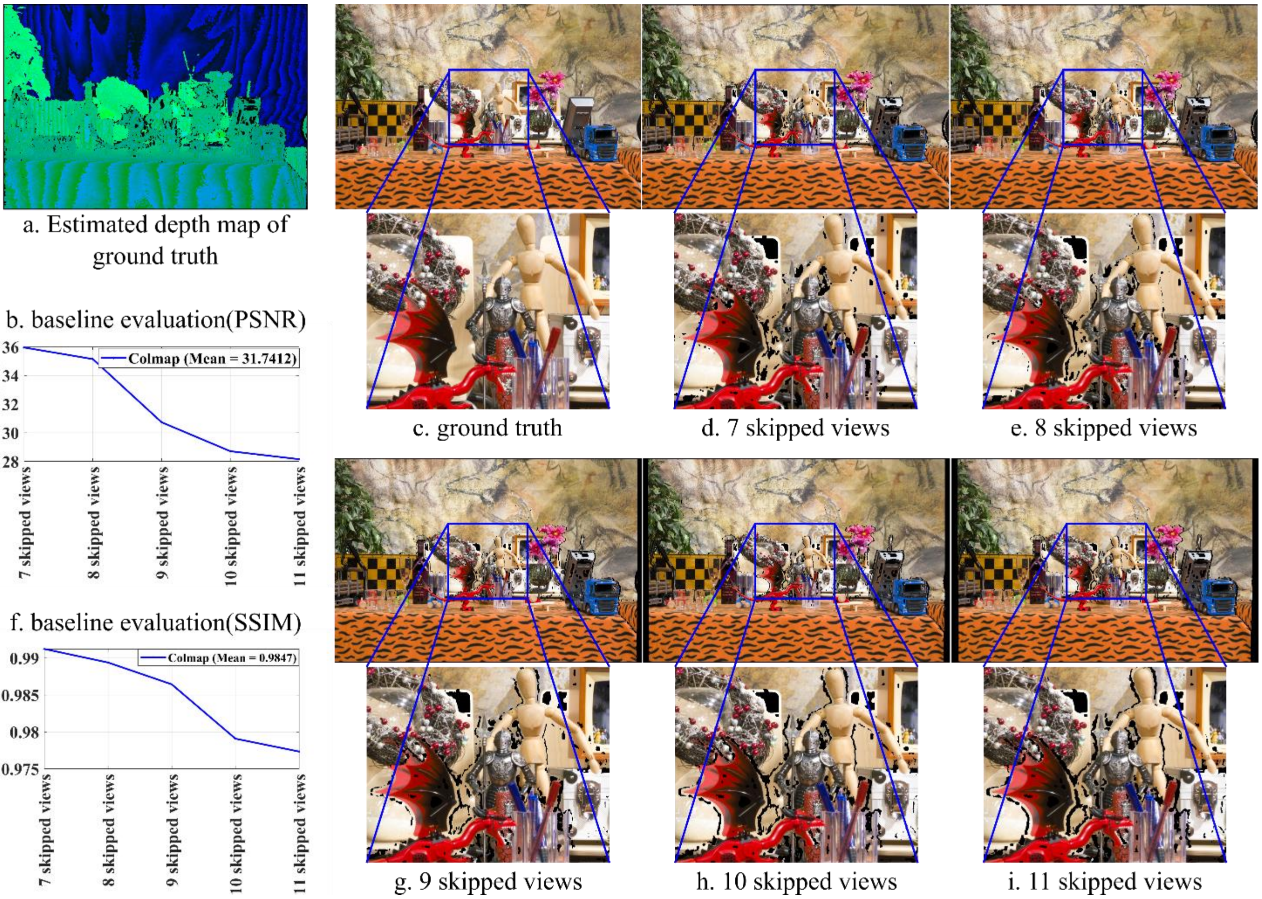


Figure 5: Non-planar concept for view rendering pipeline. The cameras filled with black color are actual cameras. The green camera is the target camera for rendering. The cameras with white color are assumed virtual partners with the baseline of 1 relative to the actual and target cameras.

For novel view rendering, we define two positions: an arbitrary target camera, and its virtual camera partner. Furthermore, we apply a forward warping for each view according to its disparity map on both of the target and its virtual partner positions. This process is done for all the 4 views within the



a. Estimated depth map of ground truth

c. ground truth

d. 7 skipped views

e. 8 skipped views

g. 9 skipped views

h. 10 skipped views

i. 11 skipped views

Figure 6: Results of reconstruction using a subset of views from *Mannequin* scene. (a). shows the estimated depth map corresponds to the ground truth. (b) and (f) show the PSNR and SSIM plots, demonstrating objective quality on the novel rendered view in comparison with the ground truth for different baselines. Visual results for different baselines are shown in (c)-(i).

neighborhood of the target position. To render a pixel on the target position, the pixel shift of the warped pixel on the target position and on the virtual partner position is calculated for each view. The calculated pixel shift reflects the distance of the object to the target camera position. Eventually, the view with the corresponding higher pixel shift wins to participate in the final rendering.

Fig. 6 shows our results using a subset of the *Mannequin* scene. The selected subset has the same density of the dataset. We first generate depth maps for our chosen subset as is shown in Fig. 6a. Then, we subsampled the dense subset by skipping views to simulate a multi-camera system with varying baselines. Each baseline is defined by skipping a fixed number of consecutive views. Moreover, the subsets, their corresponding depth maps, and the calibration information are imported into our non-planar rendering pipeline with a defined target camera position. The target position in this test is defined as the position of one of the skipped views. This skipped view later will be used as the ground truth. In the end, we rendered novel view on the same target position for each subset and then compared the novel view with the ground truth visually and objectively. By skipping views, a simulated, less dense multi-camera array is created where the baseline is larger. The visual comparison shows that a denser capture, i.e., a smaller baseline between the cameras, results in less occlusions and problems arising of that. This is a consequence of having more information of the objects from different perspectives. For the objective evaluation of the RGB information, the regions with holes are filled. Besides, the borders around the results are kept. The PSNR and SSIM results both show decreasing behavior when the baseline increases as is shown in Fig. 6b and Fig. 6f.

6. CONCLUSION

In this work, we presented a novel dense non-planar light-field dataset and rendering pipeline. The industrial robot we used for capturing, allows positioning of the RGB camera with negligible error. The dataset is further pre-processed and used for rendering novel views, analyzing the effect of a varying baseline on the image quality, and evaluating the results visually and objectively.

Other features of the dataset like high resolution images can be used for applications like VR, environmental scanning, and 3D displays. More importantly the non-planar setup of the camera positions with high accuracy in terms of capturing precision is a missing feature of the current available light-field datasets. This dataset can be used for testing in a variety of image related algorithms.

7. ACKNOWLEDGMENT

The work in this paper was funded from the European Union’s Horizon 2020 research and innovation program under the Marie

Sklodowska-Curie grant agreement No 676401, European Training Network on Full Parallax Imaging. This work is supported by the Fraunhofer and Max Planck cooperation program within the German pact for research and innovation (PFI). The capture facilities were provided by the Centre for Immersive Visual Technologies (CIVIT) at Tampere University.

8. REFERENCES

[1] G. Lippmann, “Epreuves reversibles. photographies integrals”, 1908, Comptes Rendus Academie des Sciences, vol. 146, pp. 446–451

[2] “Open Source 3D creation”, [Online]. Available: https://www.blender.org/

[3] Stanford Graphics Laboratory,“The (new) Stanford Light Field Archive”, 2008, [Online]. Available: http://lightfield.stanford.ed

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

[5] Vk. Adhikarla, M. Vinkler, D. Sumin, R. Mantuik, K. Myszkowski, H. Seidel, and P. Didy, “Towards a quality metric for dense light fields”, In Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2017, [Online]. Available: http://lightfields.mpi-inf.mpg.de/Dataset.html

[6] CIVIT lab, Tampere university of Technology, [Online]. Available: http://www.tut.fi/civit/densely-sampled-light-field-datasets/

[7] Ł. Dąbała, M. Ziegler, P. Didyk, F. Zilly, J. Keinert, K. Myszkowski, H. Seidel, P. Rokita, T. Ritschel, “Efficient Multi-image Correspondences for On-line Light Field Video Processing”, Computer Graphics Forum 35(7), In Proceedings of. Pacific Graphics 2016, DOI: 10.1111/cgf.13037, Okinawa, Japan, [Online]. Available: http://resources.mpi-inf.mpg.de/LightFieldVideo/Dataset.html

[8] N. Sabater, G. Boisson, B. Vandame, P. Kerbiriou, F. Babon, M. Hog, T. Langlois, R. Gendrot, O. Bureller, A. Schubert, V. Allie, *“*Dataset and Pipeline for Multi-View Light-Field Video*“,* In Proceedings of CVPR Workshops, 2017, [Online]. Available: *https://www.technicolor.com/dream/research-innovation/light-field-dataset*

[9] M. Ziegler, .R. Veld, J. Keinert, F. Zilly, “Acquisition System for dense lightfield of large Scenes “, In Proceedings of 3DTV Conference, Copenhagen , Denmark, DOI: 10.1109/3DTV.2017.8280412, 2017

[10] KUKA AG, https:/www.kuka.com/en-de/products/robot-systems/industrial-robots/kr-cybertech,%20kein%20Datum

[11] https://www.xrite.com/service-support/downloads/c/colorchecker_camera_calibration_v1_1_1

[12] JL. Schonberger, J. Frahm, ” Structure-From-Motion Revisited”, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, DOI: 10.1109/*CVPR*.2016.445, pp. 4104-4113.

[13] ETH Zurich, and UNC Chapel Hill, “Colmap” 2018, [Online]. Available: https://colmap.github.io

[14] R. Hartley, A. Zisserman, “Multiple view Geometry in Computer Vison”, 2nd edition, ISBN:0521540518, Cambridge University Press, New York, NY, USA, 2003.