THE PLENOPTIC 2.0 TOOLBOX: BENCHMARKING OF DEPTH ESTIMATION METHODS FOR MLA-BASED FOCUSED PLENOPTIC CAMERAS

Luca Palmieri[⋆]*, Ron Op Het Veld*[†]*, Reinhard Koch*[⋆]

[⋆]Department of Computer Science, Kiel University, Germany
[†]Department Moving Pictures Technologies, Fraunhofer IIS, Erlangen, Germany

ABSTRACT

MLA-based focused plenoptic cameras, also called type 2.0 cameras, have advantages over type 1.0 plenoptic cameras, because of their better inherent spatial image resolution and their compromise between depth of focus and angular resolution. However, they are more difficult to process since they require a depth estimation first to compute the all-in-focus image from the raw MLA image data. Current toolboxes for plenoptic cameras only support the type 1.0 cameras (like Lytro) and cannot handle type 2.0 cameras (like Raytrix). In addition, there is a lack of ground truth data and high quality benchmarking data for focussed plenoptic cameras. This contribution will discuss the requirements for processing type 2.0 images and will supply the reader with an open-source toolbox for comparing depth estimation methods. Different depth-estimation methods for MLA-based imaging will be available and an easy extension for other processing algorithms like compression will be included. In addition, we will supply benchmarking data of focused plenoptic cameras by synthetic ground truth datasets and high-quality real images captured under controlled conditions by Raytrix cameras.

Index Terms— Lightfield, Multi-focus Plenoptic, Toolbox, Dataset, MLA (Micro-lens Array)

1. INTRODUCTION

The recent growth in lightfield technologies combined with the latest research results has highlighted some major challenges within the lightfield community. From the definition itself, what is called lightfield and how different subtypes can be distinguished, to the evaluation of different approaches, where a lack of adequate material is shown.

While for binocular stereo images several different datasets covering a wide range of setup and applications have been proposed, this is not the case for lightfield data. A recent analysis of stereo vision datasets in [1] takes 28 sets into consideration, the most famous being Middlebury [2] and KITTI [3], that alone reached more than 100 citation in the three main conferences (CVPR, ICCV, ECCV) in 2016 [1].

Lightfield datasets increased in the last year and different types of images are available, like synthetic images in [4], [5], real images taken with Lytro cameras [6], [7], [8] and images taken with a moving camera or an array of cameras in [9], but they show some important limitations. Firstly, most come without ground truth, as only [5] constitutes a real benchmark for numerical comparisons, and secondly they consist of the same type of images, not taking into account all possible lightfield imagery.

None or little effort has been focused on creating a dataset for plenoptic 2.0 images. Few attempts were made in [10] and in [11] to use OpenGL to emulate camera behaviour to obtain a plenoptic image of the Stanford Bunny, but no consistent dataset was produced. Introduced in [12], multi-focus plenoptic cameras deliver another solution to the challenge of capturing lightfield in a single shot, and especially now after the discontinuation of Lytro cameras, they retain importance in the plenoptic field.

2. PRIOR WORK AND CONTRIBUTION

Many approaches have been proposed for lightfield and its several applications; it is therefore difficult and out of the scope of this work to review the whole literature. The focus will be on depth estimation, core of several applications, especially in the case of the plenoptic 2.0 cameras, because it needs geometry estimation of the scene in order to use the refocusing properties typical for lightfield acquired scenes.

Recently a combined effort of many scientists in the field was made to evaluate and analyze different depth estimation approaches using the same dataset [13]. The chosen dataset, created with a Blender plug-in to render lightfield scenes, was based on the images acquired with plenoptic 1.0 cameras and consisted of 9x9 views. Other types of lightfields were not taken into consideration.

The works focusing on the plenoptic 2.0 cameras starts with ranges from calibration, [14], [15], [16], to spatial resolution in [17] to depth estimation. Dense depth maps were achieved through stereo matching in [18], while in [19] feature matching is used to obtain sparse depth maps that are

⋆ lpa@informatik.uni-kiel.de
† The author performed the work while at Kiel University
The work in this paper was funded from the European Unions Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 676401, European Training Network on Full Parallax Imaging.

successively filled. Recent works focus on optimization or mixing of lens pattern selections and the creation of synthetic images, either with Blender or OpenGl, as in [20] and [11].

Therefore our work provides a tool for multi-focus plenoptic cameras to contribute to the research of different lightfield acquisition methodologies.

The contribution of this work is twofold:

1. It provides a toolbox for plenoptic 2.0 images, including a dataset of such images, consisting of both real images taken with Raytrix cameras (where raw and processed images are available along with a configuration file) and synthetic images with relative ground truth and a completely open-source repository with the code used to work on these images;

2. It evaluates different methods to estimate depth from these kind of images using both synthethic numerical evaluation purposes and real images.

3. THE PLENOPTIC TOOLBOX

The Plenoptic Toolbox consists of open-source code that allows the development of several applications using plenoptic 2.0 images. All the source code is available online at the GitHub page [21], in the *python* language, for research purposes.

The provided code uses a dictionary to load and store the micro-images with their parameters. This allows for an efficient utilization and offers the flexibility to develop new methods and applications without the need of a new implementations.

3.1. The Dataset

Given the challenges related to the acquisition and the usage of multi-focus plenoptic cameras, few reliable images are available online: our work provides a complete online database of plenoptic 2.0 images, taken with different cameras to guarantee a homogeneous distribution.

At the time of this publication, Raytrix R29 and R42 cameras were chosen because of their higher quality of the pictures. All pictures were taken under controlled conditions.

The creation of the dataset followed the plan to present different challenges in the depth estimation: as it's possible to see in the supplementary material, the acquired images present white background and thus textureless regions as in *Cards* or *Cars*, as well as textured background in *University* or *Dragon*, different types of specularities in *Specular*, fine and detailed structures in *Hawaii*, and slopes with texture for an easier matching in *Dixit*.

As visible in Fig. 1, a set of synthetic images is provided along with the real images: they were created using Blender to emulate the micro-lens array grid, therefore in an ideal condition. No distortion have been applied, and they all have the corresponding ground truth.

The multi-focus properties of the cameras are also taken into account in the generation of the synthtetic images, delivering micro-images with different amount of blur according to their focal lengths, to be as close as possible to the real scenes.

4. BENCHMARKING OF DEPTH ESTIMATION

Due to their recent introduction, there are not many technique forming the state-of-the-art for depth estimation on plenoptic 2.0 images. As a start, work from [20] and [18] is used, where depth is computed by stereo matching the micro-images using the well known semi-global matching method.

Five different similarity measures to compute the cost volume are analyzed: Absolute Difference (AD), Squared Difference (SD), Census, Gradient with AD and Normalized Cross Correlation. These measures are widely used and constitutes the basis for state-of-the-art methods, as in [23] where a learning approach is used to choose the best matching costs.

Similar evaluations refer to binocular stereo [24] and do not account for the multi-view case or plenoptic micro-images, thus our evaluation extends these works to plenoptic images.

4.1. Evaluation

The evaluation procedure is divided into two parts, because of the nature of the different images. For the real images, only a visual subjective comparison can be made, as already highlighted in [19], due to the lack of ground truth. We provide some visual comparison between the different methods (more in the supplementary material) and a general overview.

In the case of synthetic images it's possible to analyze more in detail the results. Following the Middlebury stereo vision benchmark [2] and the latest works in this field as [5], [13] and [25], the most representative criteria for depth estimation classification has been reproduced.

The criteria to be used for benchmarking were chosen based on their significance. For example, with respect to binocular stereo vision, plenoptic 2.0 images present special properties: such images are less affected by the occlusion problem, so this criterium, of large importance in the stereo case, was discarded for our benchmark.

The errors have been calculated under the form of:

- Number of pixels that show a disparity error larger than ϵ, with $\epsilon \in [0, 2]$ are estimated. As a reference the two values for $\epsilon = 1, 2$ are chosen, like the socalled *bad 1.0* and *bad 2.0* of Middlebury).

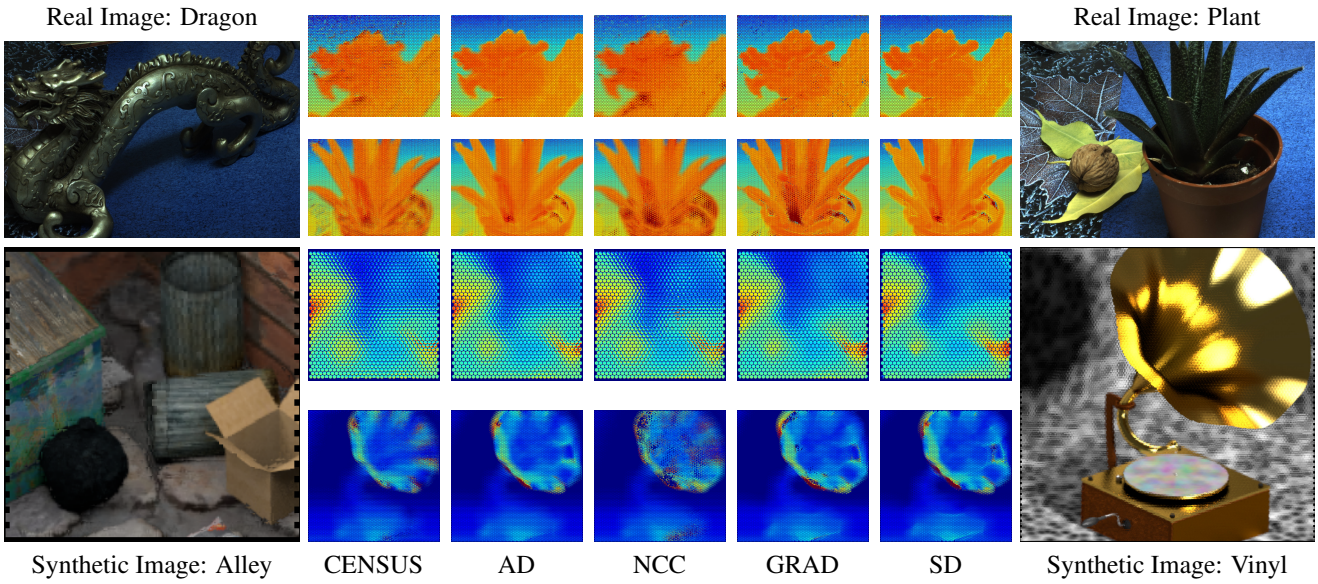

Fig. 1. Samples of images from the dataset with their respective estimated depth maps. *First row:* real images taken with R29 Raytrix camera, along with an excerpt of their depth map estimated using the similarity measures above described. *Second row:* synthetic images generated with Blender, along with their depth maps. The ground truth is not shown here. *AD* = absolute difference. *NCC* = normalized cross correlation. *GRAD* = gradient. *SD* = squared difference. Please refer to the colored version for a better visualization. Images are visible in the supplementary material and available at [22].

- Average Error (AE) and Mean Squared Error (MSE)

$$AE = \sum_{i,j \in I} \frac{|d_{i,j} - gt_{i,j}|}{|I|} \quad MSE = \sum_{i,j \in I} \frac{(d_{i,j} - gt_{i,j})^2}{|I|} \quad (1)$$

- Bumpiness measurement, that accounts for smoothness of estimation, using the formula from [5], changing the clamping threshold (B_{thresh}) according to our depth range.

$$B = \sum_{i,j \in I} \frac{\min(B_{thresh}, |d_{i,j} - gt_{i,j}|)}{|I|} \quad (2)$$

  where in the presented results $B_{thresh} = 0.25$, $d_{i,j}$ and $g_{i,j}$ represent respectively disparity estimated and ground truth at the i and j pixel, and I indicates the circle-shaped micro-image that is used for the calculations.

- Errors around depth discontinuities, similar to the criterium used in [25]. The image is divided into two parts, where one contains the pixels around edges in the disparity maps and the other one the rest. The edges were obtained through the OpenCV implementation of the Canny algorithm followed by a dilation operation to obtain the area around it (one pixel per side).

4.2. Results on Real Images

In this section some of the results are shown to back up the general considerations. Because of the limitations of synthetic images, real images still provide for a more challenging task, having to deal with physical lenses and sensors. Moreover, the variety of the scenes that can be captured allow the testing of the algorithms for different purposes, targeting specific issues.

The results show that the current algorithm, independently from the similarity measure chosen, fails in reconstructing large textureless surfaces. For this reason, two different sets of images were acquired: one with white textureless background, where the algorithm shows low quality results, and one with textured background, where it delivers robust estimation. This happens because of the local approach used and can be improved by choosing a global solution or by applying a post processing refinement step, for example a filling algorithm. This will be addressed in future research.

The similarity measures analyzed show quite some difference in the analyzed images. The AD and CENSUS obtain satisfactory results and high quality depths in textured scenes, while NCC shows alternate performances depending on the scene and large error that affect the whole image, so that its usage can be considered mainly in combination with other simlarity measure. The depth maps obtained with SD and Gradient, lastly, show lower quality and larger areas with a wrong estimation.

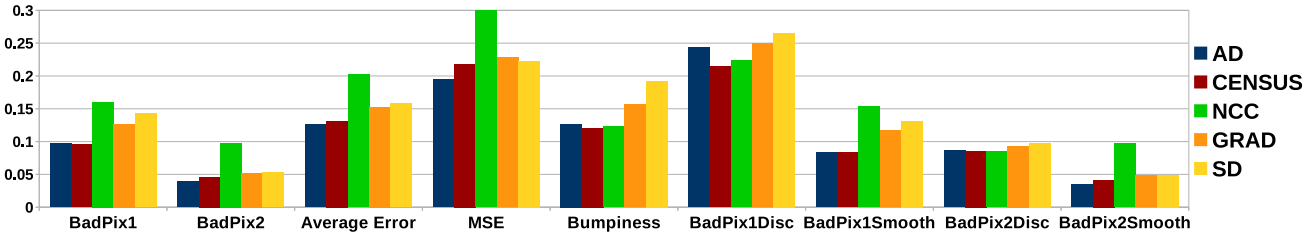

Fig. 2. The different criteria used for evaluation. Some of the values have been scaled for visualization purposes. *BadPix1,2* = Percentage of pixels which error exceeds 1,2 pixels. *Average Error* = Average of the absolute error in pixel. *MSE* = Mean squared Error. *Bumpiness* = Bumpiness measure taken from [5]. *..Disc* = .. Around depth discontinuities. *..Smooth* = .. Around smooth areas (not considered depth discontinuities.) Please refer to the colored version for a better visualization.

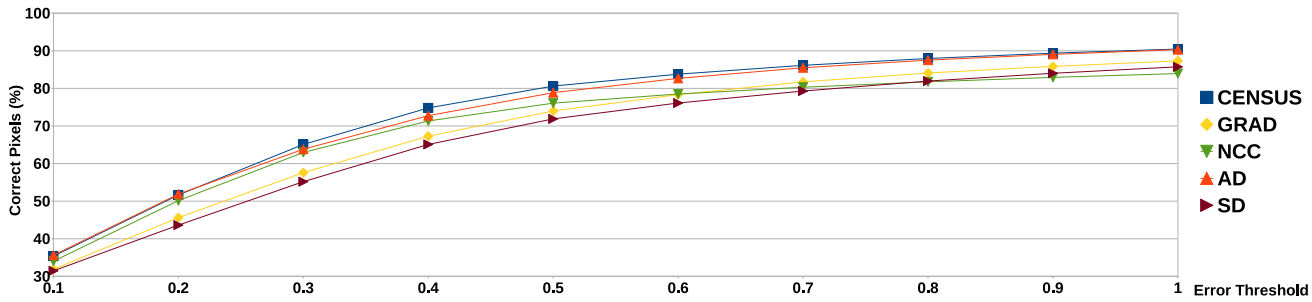

Fig. 3. For each similarity measure, the percentage of correctly estimated pixel on the synthetic scenes is plotted for the increasing error's thresholds on the x-axis. Please refer to the colored version for a better visualization.

4.3. Results on Synthetic Images

The synthetic images were analyzed on the above defined criteria: AD and CENSUS are the methods who achieve an overall higher quality, with NCC that shows some interesting characteristics and Gradient and SD that fail larger areas. On the average error measure and the Bad Pixel 1.0 / 2.0, the AD and CENSUS outperform the other methods and achieve comparable results. The same is visible for the Mean Squared Error, apart from a higher value for NCC. NCC has lower performances in the Bad Pixel 1.0 and 2.0, showing the highest number of errors with large value. In the Bumpiness criterium AD, CENSUS and NCC reach the same level and SD and GRAD have weaker performances.

The last criteria, indicating the number of erroneous pixels around depth discontinuities and in smoother areas, give a reference about the robustness of the estimation: NCC, for example, obtains a good score in the discontinuities areas, while performing poorly on smooth surfaces. Apart from the NCC case, the errors confirm that depth discontinuities are still the most challenging parts. This might be emphasized by our approach that does not include a refinement step for accurate reconstruction of fine structures.

Fig. 3 shows the behaviour of the number of correct pixels varying the error threshold. As expected, AD and CENSUS reach higher levels. The estimation using NCC has an unexpected curve: for low thresholds it maintains same level of AD and CENSUS, but makes higher amount of larger errors, resulting to lower performances in most of the measurements. This suggests that a combination of NCC with other measurements based on its confidence could lead to improvements.

5. CONCLUSION

The presented work extends the benchmarking for stereo vision to the plenoptic 2.0 images. It contributes to the development and spreading of such technologies, providing an important tool for future research; moreover, it makes available a dataset of images that is, up to our knowledge, the first of its kind. The dataset is meant to be continuously updated with new images for specific purposes and increasing difficulties.

This will not only allow an easier comparison of different methodologies of disparity estimation techniques, but also push other possible applications: in this direction, the next step will be to develop novel approaches for compression of such images.

Lastly, we provide a first version of plenoptic 2.0 benchmark, where different similarity measures are compared. Even though at the current state-of-the-art the changes are quite simple, it is to be seen as standard for an easy comparison for future development, a missing tool in this field.

6. REFERENCES

[1] O. Zendel, K. Honauer, M. Murschitz, M. Humenberger, and G. F. Domìnguez, "Analyzing computer vision data - the good, the bad and the ugly," in *Computer Vision and Pattern Recognition (CVPR)*, 2017.

[2] D. Scharstein, R. Szeliski, and H. Hirschmuller, "Middlebury stereo vision benchmark," `http://vision.middlebury.edu/stereo/`, [Online, 2018].

[3] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Computer Vision and Pattern Recognition*, 2012.

[4] G. Wetzstein, "Synthetic light field archive," `http://web.media.mit.edu/~gordonw/SyntheticLightFields/index.php`, [Online, 2018].

[5] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldluecke, "A dataset and evaluation methodology for depth estimation on 4d light fields," in *Asian Conference on Computer Vision (ACCV)*, 2016.

[6] A. Mousnier, E. Vural, and C. Guillemot, "Lytro first generation dataset," `https://www.irisa.fr/temics/demos/lightField/index.html`, [Online, 2018].

[7] A. Ghasemi, N. Afonso, and M. Vetterli, "Lcav-31: A dataset for light field object recognition," in *Proceedings of the SPIE, vol. 9020, International Society for Optics and Photonics*, 2014.

[8] Caner Hazirbas, "4.5d lightfield-depth benchmark," `http://hazirbas.com/datasets/ddff12scene/`, [Online, 2018].

[9] A. S. Raj, M. Lowney, and R. Shah, "Light-field database creation and depth estimation," `http://lightfield.stanford.edu/`, 2016.

[10] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *IEEE International Conference on Computational Photography (ICCP)*, 2009.

[11] R. Ferreira, J. Cunha, and N. Goncalves, "Multi-focus plenoptic simulator and lens pattern mixing for dense depth map estimation," in *EUROGRAPHICS*, 2016.

[12] C. Perwaß and L. Wietzke, "Single lens 3d-camera with extended depth-of-field," in *Proceedings of SPIE - The International Society for Optical Engineering 8291:4*, 2012.

[13] K. Honauer and O. Johannsen et al., "A taxonomy and evaluation of dense light field depth estimation algorithms," in *Computer Vision and Pattern Recognition (CVPR)*, 2017.

[14] S. Nousias, F. Chadebecq, J. Pichat, P. Keane, S. Ourselin, and C. Bergeles, "Corner-based geometric calibration of multi-focus plenoptic cameras," in *International Conference on Computer Vision (ICCV)*, 2017.

[15] Y. Bok, H.G. Jeon, and I.S. Kweon, "Geometric calibration of micro-lens-based light field cameras using line features," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, February 2017.

[16] C. Heinze, S. Spyropoulos, S. Hussmann, and C. Perwaß, "Automated robust metric calibration algorithm for multifocus plenoptic cameras," *IEEE Transactions on Instrumentation and Measurements*, May 2016.

[17] M. Damghanian, R. Olsson, M. Sjöström, A. Erdmann, and C. Perwaß, "Spatial resolution in a multi-focus plenoptic camera," in *International Conference on Image Processing (ICIP)*, 2014.

[18] O. Fleischmann and R. Koch, *Lens-Based Depth Estimation for Multi-focus Plenoptic Cameras*, pp. 410–420, Springer International Publishing, 2014.

[19] R. Ferreira and N. Goncalves, "Fast and accurate micro lenses depth maps for multi-focus light field cameras," in *German Conference on Pattern Recognition*, 2016.

[20] L. Palmieri and R. Koch, "Optimizing the lens selection process for multi-focus plenoptic cameras and numerical evaluation," in *2nd LF4CV Workshop*. CVPR, 2017.

[21] L.Palmieri, "The plenoptic toolbox 2.0," `https://github.com/PlenopticToolbox/PlenopticToolbox2.0`, [Online, 2018].

[22] L. Palmieri, "Multi-focus plenoptic images dataset," `http://dx.doi.org/10.17632/t6czryg5nw.1`, [Online, 2018].

[23] H. G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y. W. Tai, and I. S. Kweon, "Depth from a light field image with learning-based matching costs," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2018.

[24] H. Hirschmuller and D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 31, no. 9, pp. 1582–1599, 2009.

[25] K. Honauer, L. Maier-Hein, and D. Kondermann, "The hci stereo metrics: Geometry-aware performance analysis of stereo algorithms," in *International Conference on Computer Vision (ICCV)*, 2015.