Downloaded from the open archive of Mid Sweden University DiVA.

This is the accepted version of a paper published in 2019 IEEE International Conference on Image Processing (ICIP). See full citation below.

Li, Yongwei; Sjöström, Mårten, "Depth-Assisted Demosaicing for Light Field Data in Layered Object Space," *2019 IEEE International Conference on Image Processing (ICIP)*, pp.3746-3750, 22-25 September 2019.
<http://dx.doi.org/10.1109/ICC.2012.6364802>

© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

DEPTH-ASSISTED DEMOSAICING FOR LIGHT FIELD DATA IN LAYERED OBJECT SPACE

Yongwei Li, Mårten Sjöström

Mid Sweden University
Department of Information Systems and Technology
SE-851 70 Sundsvall Sweden

ABSTRACT

Light field technology, which emerged as a solution to the increasing demands of visually immersive experience, has shown its extraordinary potential for scene content representation and reconstruction. Unlike conventional photography that maps the 3D scenery onto a 2D plane by a projective transformation, light field preserves both the spatial and angular information, enabling further processing steps such as computational refocusing and image-based rendering. However, there are still gaps that have been barely studied, such as the light field demosaicing process. In this paper, we propose a depth-assisted demosaicing method for light field data. First, we exploit the sampling geometry of the light field data with respect to the scene content using the ray-tracing technique and develop a sampling model of light field capture. Then we carry out the demosaicing process in a layered object space with object-space sampling adjacencies rather than pixel placement. Finally, we compare our results with state-of-art approaches and discuss about the potential research directions of the proposed sampling model to show the significance of our approach.

Index Terms— Light field, demosaicing, object space, ray-tracing technique

1. INTRODUCTION

Preserving the high-fidelity images of a scene is a challenging task. A conventional camera maps three-dimensional objects onto a two-dimensional planar sensor by a projective transformation. Based on such a 2D image, depth perception can only be approximated by the sophistication of empirical human visual system (HVS), or considerably complex computer vision algorithms [1]. In order to fully reconstruct the scene, the plenoptic function [2] is required. Stemming from the concept of plenoptic function, Levoy et al. [3] and Gortler et al. [4] further proposed an equivalent two-plane representation of the plenoptic function, referred to as the light field or

the lumigraph.

The 4D light field can be captured with various setups, such as the multi-camera rig and the plenoptic camera. A multi-camera rig uses conventional cameras to capture different perspectives of a scene with a large baseline [3], and plenoptic cameras employ microlenses to decouple the spatial and angular information [5, 6]. Both the multi-camera rig and plenoptic cameras capture a sampled light field, enabling applications which cannot be achieved with a single view, such as depth estimation [7, 8] and refocusing [6, 8]. However, the problems in terms of the raw data processing has been barely discussed, given the fact that ill-suited pre-processing steps will propagate erroneous information through the entire follow-up workflow [9, 10].

In this work, we focus on the demosaicing problem of light field data and propose a generic demosaicing framework. Compared with conventional demosaicing methods [11] which implicitly assume the scene sampling consecutiveness in neighboring pixels, our demosaicing scheme works directly in the object space by projecting pixels back to 3D space. In other words, we demosaic the vicinity in the object space rather than the pixel adjacency on a sensor. By doing so, our work introduces a more suitable approach based on the sampling property of the cameras, yielding an accurate, physically-based color interpolation.

The remainder of this paper is organized as follows: we first review the state-of-art light field demosaicing methods in Section 2. Section 3 describes our demosaicing scheme in detail. The demosaicing comparison with state-of-art methods and further analysis can be found in Section 4. Finally, in Section 5, we conclude this paper and indicate possible future research.

2. LIGHT FIELD DEMOSAICING

In a color image, there is a unique set of RGB values at each pixel location. To combat the conflict between a single CCD sensor and multiple color channels, most digital cameras are equipped with a color filter array (CFA). In order to make the best estimate to fill in the remaining two channels to each pixel, a demosaicing process is required. The demosaicing problem for conventional cameras has been widely discussed

The work in this paper was funded from the European Unions Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 676401, European Training Network on Full Parallax Imaging.

for years, including approaches both in spatial and frequency domain [11]. Classic demosaicing methods aim to interpolate between nearby samples on the sensor. This is due to the fact that the spatial sampling frequency of the scene is determined by the pixel pitch for conventional cameras. However, the above rule no longer holds for light field captures, considering the discordance between the vicinity on the sensor and that in the sampling space [12].

An intuitive way to demosaic the plenoptic image is to interpolate from the vicinity for any missing channel as described in a widely used pipeline [13], which handles the crosstalk effects by ignoring the pixels near the lenslet edges. Based on this, David et al. [14] further improved the performance by using a white elemental image to overcome the vignetting problem and weight the interpolation within the elemental image. Yu et al. [15] proposed to resample the light field based on the refocusing and postpone the demosaic process until the rendering stage. However, it aims at demosaicing the refocused images and the performance varies depending on the refocusing plane. Georgiev et al. [12] analyzed the sampling of adjacent elemental images and proposed to interleave each color component from nearest elemental images before demosaicing, such that superresolution can be achieved in joint with demosaicing by a small effort. A learning based demosaicing method was proposed in [16], which uses K-SVD algorithm to learn from a set of reordered sample vectors of the light field. Then the demosaicing is done by solving the sparse reconstruction problem of the system. However, none of the aforementioned methods explicitly consider the impact of scene depth in the demosaicing. The only exception is [17], where several claims are made by analyzing a sampling model, without any specific demosaicing approach.

3. OBJECT SPACE DEMOSAICING

Most of the existing demosaicing schemes introduced in Section 2 are performed on condition that each elemental image captures the identical spatial information from various angles. This inference is true if the decoded subaperture images have a considerably small baseline and the scene lies on the focal plane. However, such schemes still suffer from visible artifacts, especially around the edges [15] in practical cases. In this section, we propose an object-space light field demosaicing method, which explores the implicit connection between the pixel sampling and the 3D scene point by ray tracing. Then we shift the demosaicing process from 2D planar sensor to a layered 3D object space so that color is recovered in the object space with physically nearest sampled spatial points. Note that although our method is generic for any light field data, the following analysis is mainly based on plenoptic cameras.

Fig. 1: A sampling model of plenoptic cameras based on the chief rays and pinhole assumption. Only rays through two microlenses are illustrated for concision. Note the changes in the light field sampling between focal plane and an arbitrary sampling plane.

3.1. Ray model of plenoptic capture

In this section, we provide a theoretical analysis of plenoptic camera sampling based on ray tracing. We model the microlens array (MLA) as a set of pinhole cameras and consider only the chief rays passing through the centers of them, as shown in Fig. 1. Note that only one dimension of the two-plane light field parameterization is depicted in the figure to simplify the illustration without losing the generality, the other dimension can be modeled in an identical manner.

It can be seen from Fig. 1 that each microlens gathers angular information of a spatial point converged by the main lens when the scene is on the focal plane. In spatial domain, the sampling period p is proportional to the microlens pitch d . Additionally, the spatial resolution of the scene is correlated with the number of microlenses N , as a subaperture image can be rendered by collecting corresponding pixels under each microlens, representing the scene from one angle of view. However, the sampling period p' becomes irregular and overlapping in the object space when the scene is not on the focal plane. In this case, each microlens samples a small region of the scene, such as AC in Fig. 1, instead of a singular spatial point. The upper bound of the sampling frequency is limited by the projection of non-adjacent pixels AC' , rather than that of adjacent sensor pixels AB . Therefore, conventional sensor-based demosaicing methods which overlook the depth are suboptimal solutions, resulting in a smear of color in unfocused planes.

Formally, we use a ray transfer matrix (RTM) to retrieve the pixel sampling pattern in the object space. We define a point on the ray at distance Z from the sensor as a vector $\mathbf{P}_{(Z)} = (X, Y, \theta, \phi, \mathbf{C}^T)^T$, where (X, Y) is the intersection point of the light ray and the depth plane Z , θ and ϕ are the angles with respect to X and Y axis, a 3-dimensional vector $\mathbf{C} = (R, G, B)^T$ is used to describe the RGB color compo-

nents.

Then we exploit the 3D RTM for ray tracing. The free space propagation from plane $\Pi: Z = z$ to $\Pi: Z = z'$ is computed as $\mathbf{P}_{(z')} = \mathbf{S}\mathbf{P}_{(z)}$, with the propagation matrix $\mathbf{S}$:

$$\mathbf{S} = \begin{pmatrix} 1 & 0 & z' - z & \mathbf{0}_4^T \\ 0 & 1 & 0 & z' - z & \mathbf{0}_3^T \\ \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{E}_5 & \mathbf{0}_3^T \end{pmatrix}, \quad (1)$$

where $\mathbf{E}$ and $\mathbf{0}$ are the identity matrices and zero vectors respectively. Similarly, the thin lens refraction matrix $\mathbf{L}$ can be written as:

$$\mathbf{L} = \begin{pmatrix} 1 & \mathbf{0}_6^T \\ 0 & 1 & \mathbf{0}_5^T \\ -\frac{1}{f} & 0 & 1 & \mathbf{0}_4^T \\ 0 & -\frac{1}{f} & 0 & 1 & \mathbf{0}_3^T \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{E}_3 \end{pmatrix}, \quad (2)$$

where f is the focal length of the lens. Finally, for an arbitrary color channel $c \in \{R, G, B\}$, the CFA is defined as three diagonal matrices: $\mathbf{F}_R = \text{diag}(0, 0, 0, 0, 0, 1, 0, 0)$, $\mathbf{F}_G = \text{diag}(0, 0, 0, 0, 0, 0, 1, 0)$, $\mathbf{F}_B = \text{diag}(0, 0, 0, 0, 0, 0, 0, 1)$, one for each pixel location based on the mosaic filter pattern. MLA is modeled as a set of pinhole cameras, passing only the chief rays while blocking the peripheral ones, given by the matrix $\mathbf{O}$:

$$\mathbf{O} = \begin{cases} \mathbf{0}_{7 \times 7}, & \text{if } (X, Y) \neq (X_l, Y_l), \\ \mathbf{E}_7, & \text{otherwise,} \end{cases} \quad (3)$$

where (X_l, Y_l) is the optical center of the corresponding lens. By incorporating all the RTMs and filtering process, a correlation between pixel and object space sampling can be obtained by matrix multiplication:

$$\mathbf{P}' = \mathbf{S}_3 \mathbf{L}_{ml} \mathbf{S}_2 \mathbf{O} \mathbf{S}_1 \mathbf{F}_{R,G,B} \mathbf{P}, \quad (4)$$

where $\mathbf{P}'$ is the sampling point in the object space for a ray emitted from pixel location $\mathbf{P}$ on the sensor. $\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3$ represent the propagation between sensor, MLA, main lens and the sampled scene, one step at a time. $\mathbf{L}_{ml}$ is the main lens refraction matrix from (2). Here we address MLA and main lens in different manners to reduce computational complexity while preserving the applicability of our model when additional main lens effects need to be considered.

3.2. Layered object space demosaicing

Given the parameters and geometrical structure of a plenoptic camera, the only unknown in the model (4) is the depth of the scene. Previous studies such as [10] has shown that a crude disparity map (Fig. 2) can be acquired before the demosaicing process by block-matching identical color channels and performing a coherence check. One can further improve the accuracy of the disparity by state-of-art correspondence searching functions in computer vision [9]. However, our goal is not

Fig. 2: Initial depth estimation for light field data. (a) The ground truth disparity map. (b) The estimated disparity map using [10] from the raw data.

to address the depth estimation problem, we instead improve the demosaicing results using the estimated depth.

In general, interpolation for missing color channels in 3D space is a complicated problem due to the non-uniform sampling on different depths. However, there exists extreme cases which are simple to cope with: 1) when the scene is entirely on the focal plane and 2) when the scene is planar and perpendicular to the optical axis of the camera, as shown in Fig. 1. In the first case, no interpolation is needed if the scene is Lambertian, as pixels of different color channels within the same elemental image capture the same scene content. Thus, color information of a 3D point $\mathbf{P}$ can be restored as:

$$I_{c,\mathbf{P}} = \frac{1}{m} \sum_{i=1}^m I_{c,i}, \quad (5)$$

where $I_{c,\mathbf{P}}$ is the averaged intensity in color channel c from m nearest points of $\mathbf{P}$. In the second case, conventional 2D interpolation can be applied. Note that even though the scene is planar, the sampling grid is irregular and the color distribution varies based on the depth. Therefore, an inverse distance weighting (IDW) function is adopted as the color interpolation method for each channel:

$$I_{c,\mathbf{P}} = \frac{\sum_{i=1}^m \omega_{c,i} I_{c,i}}{\sum_{i=1}^m \omega_{c,i}}, \quad (6)$$

where $\omega_{c,i} = (d_{c,i})^{-n}$ is the weighted distance between $\mathbf{P}$ and its i -th nearest point of color c , and n defines the decaying of weighting with the distance $d_{c,i}$ which affects the smoothness of the interpolation.

Except for the cases which are discussed above, one has to solve the intractable problem of interpolation from irregular samples in 3D space. In addition, depth estimation generates defective results as shown in Fig. 2, which have a negative impact on locating nearest 3D samples when demosaicing in the object space. To jointly solve these problems, we propose to demosaic the light field in a layered object space scheme

Fig. 3: Demosaicing results using different algorithms: (a) [13], (b) [14] and (c) Proposed layered object space demosaicing. The zoomed regions are shown to the right.

such that we can achieve desired results with low computational effort and high robustness to the depth uncertainty.

We first divide the object space into $q \geq 2$ layers between a defined minimum depth $z_{\min}$ and maximum depth $z_{\max}$, so that each layer contains $(q - 1)^{-1}$ of the depth range $z_{\max} - z_{\min}$. For any sampled point (x, y, z) in the object space, the

closest depth layer Π : $Z = z'$ can be determined by:

$$z' = \left\lceil \frac{z}{\Delta z} - \frac{1}{2} \right\rceil \cdot \Delta z, \quad (7)$$

where $\Delta z = \frac{z_{\max} - z_{\min}}{q - 1}$ is the bin size of equally distributed depth range for one layer, and $\lceil \cdot \rceil$ is the ceiling function. Then the sampling points are projected to the closest layer by finding the intersection $\mathbf{P}'$ of a ray with the plane Π : $Z = z'$ using (1). By far, we mapped sampled points in the object space to a set of planes, resulting in a sequence of extreme situations described in the second case above. Thus, (6) can be employed to solve the interpolation problem on q planes instead of coping with irregular 3D samples.

4. SIMULATION RESULTS AND DISCUSSION

As stated in the previous section, our demosaicing method can be extended to any form of light field capturing devices, such as multi-camera or movable camera gantry, with a small effort: change the main lens focal length f to infinity in (2). One should note that in case of large baseline, the scene is sparsely sampled and it is necessary to constrain the distance parameter $d_{c,i}$ in (6) within a reasonable interval.

We perform our experiments with a preliminary estimated depth as shown in Fig. 2b. It can be seen that the depth information is stable within the contour of the same object, and the inaccurate depth mostly exists in poorly textured areas. However, such homogeneous regions contain only low-frequency information, which makes the interpolation less sensitive to depth uncertainty. Essentially, q sets a trade-off between robustness and correctness of demosaicing. In case of an accurate depth estimation, a large q (fine depth resolution) implies a more accurate demosaicing. On the other hand, having a poor depth estimation, a small q can be more advantageous to ensure the sampling adjacency.

The final result of the layered object space demosaicing scheme is shown in Fig. 3, compared with the reference demosaicing methods [13] and [14]. The zoomed figures show that the basketball are filled with noisy information for references, while our approach preserves more sharp results. The shape of red flowers on wallpaper is visually perceivable from our method, while the other two methods fail to reconstruct it.

5. CONCLUSION

In this paper, we proposed a simple and effective object space demosaicing scheme for light field images. We circumvented the edge detection problem in conventional demosaicing and prevented crosstalk artifacts by a layered approach. So far, the experimentation is carried out with simplified assumptions, such as pinhole, thin lens and Lambertian scene. The effects of lens aberrations and non-Lambertian scenes are left as future work.

6. REFERENCES

- [1] Ashutosh Saxena, Min Sun, and Andrew Y Ng, “Make3d: Learning 3d scene structure from a single still image,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 5, pp. 824–840, 2009.
- [2] James R Bergen and Edward H Adelson, “The plenoptic function and the elements of early vision,” *Computational Models of Visual Processing*, pp. 3–20, 1991.
- [3] Marc Levoy and Pat Hanrahan, “Light field rendering,” in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 31–42.
- [4] Steven J Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F Cohen, “The lumigraph,” in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 43–54.
- [5] Andrew Lumsdaine and Todor Georgiev, “The focused plenoptic camera,” in *Computational Photography (ICCP), 2009 IEEE International Conference on*. IEEE, 2009, pp. 1–8.
- [6] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan, “Light field photography with a hand-held plenoptic camera,” *Computer Science Technical Report CSTR*, vol. 2, no. 11, pp. 1–11, 2005.
- [7] Ting-Chun Wang, Alexei A Efros, and Ravi Ramamoorthi, “Occlusion-aware depth estimation using light-field cameras,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3487–3495.
- [8] Yongwei Li, Gabriele Scrofani, Mårten Sjöström, and Manuel Martinez-Corral, “Area-based depth estimation for monochromatic feature-sparse orthographic capture,” in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 206–210.
- [9] Mozhdeh Seifi, Neus Sabater, Valter Drazic, and Patrick Perez, “Disparity-guided demosaicking of light field images,” in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 5482–5486.
- [10] Neus Sabater, Valter Drazic, Mozhdeh Seifi, Gustavo Sandri, and Patrick Pérez, “Light-field demultiplexing and disparity estimation,” in *technical report: hal-00925652*, 2014.
- [11] Xin Li, Bahadır Gunturk, and Lei Zhang, “Image demosaicing: A systematic survey,” in *Visual Communications and Image Processing 2008*. International Society for Optics and Photonics, 2008, vol. 6822, p. 68221J.
- [12] Todor Georgiev, Georgi Chunev, and Andrew Lumsdaine, “Superresolution with the focused plenoptic camera,” in *Computational Imaging IX*. International Society for Optics and Photonics, 2011, vol. 7873, p. 78730X.
- [13] Donald G Dansereau, Oscar Pizarro, and Stefan B Williams, “Decoding, calibration and rectification for lenselet-based plenoptic cameras,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 1027–1034.
- [14] Pierre David, Mikaël Le Pendu, and Christine Guillemot, “White lenslet image guided demosaicing for plenoptic cameras,” in *Multimedia Signal Processing (MMSP), 2017 IEEE 19th International Workshop on*. IEEE, 2017, pp. 1–6.
- [15] Zhan Yu, Jingyi Yu, Andrew Lumsdaine, and Todor Georgiev, “An analysis of color demosaicing in plenoptic cameras,” in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 901–908.
- [16] Xiang Huang and Oliver Cossairt, “Dictionary learning based color demosaicing for plenoptic cameras,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 449–454.
- [17] Yongwei Li, Roger Olsson, and Mårten Sjöström, “An analysis of demosaicing for plenoptic capture based on ray optics,” in *3D at any scale and any perspective, 3-5 June 2018, Stockholm–Helsinki–Stockholm*, 2018.