![DiVA logo](http://www.diva-portal.org)
This is the published version of a paper presented at *3D at any scale and any perspective, 3-5 June 2018, Stockholm – Helsinki – Stockholm.*

N.B. When citing this work, cite the original published paper.

AN ANALYSIS OF DEMOSAICING FOR PLENOPTIC CAPTURE BASED ON RAY OPTICS

Yongwei Li, Roger Olsson, Mårten Sjöström

Department of Information Systems and Technology, Mid Sweden University
Sundsvall, Sweden SE-85170

ABSTRACT

The plenoptic camera is gaining more and more attention as it captures the 4D light field of a scene with a single shot and enables a wide range of post-processing applications. However, the pre-processing steps for captured raw data, such as demosaicing, have been overlooked. Most existing decoding pipelines for plenoptic cameras still apply demosaicing schemes which are developed for conventional cameras. In this paper, we analyze the sampling pattern of microlens-based plenoptic cameras by ray-tracing techniques and ray phase space analysis. The goal of this work is to demonstrate guidelines and principles for demosaicing the plenoptic captures by taking the unique microlens array design into account. We show that the sampling of the plenoptic camera behaves differently from that of a conventional camera and the desired demosaicing scheme is depth-dependent.

Index Terms — Light field, plenoptic camera, depth, image demosaicing

1. INTRODUCTION

Since Lippmann [1] first proposed integral photography (IP), tremendous efforts have been made in capturing and recreating 3D scenes. A notable advance towards this goal is the plenoptic camera, which utilizes a microlens array (MLA) to decouple the spatial and angular information on a image sensor. Such design enables several post-processing applications, ranging from depth estimation [2] to super-resolution [3]. Different types of plenoptic cameras have been reported during the last decades [4, 5].

While a large amount of work has been done in improving the resolution of plenoptic cameras [3, 6, 7], demosaicing has not been profoundly studied. Plenoptic cameras, such as Lytro Illum [4], capture color information by placing a color filter array (CFA) in front of the sensor in the same way as a conventional camera. Each pixel of the recorded raw sensor image collects either red, green or blue information. In order to restore the full-resolution color image, demosaicing is applied to make the best estimate and fill in the remaining other two channels for each pixel [8].

In this paper, we discuss the depth-dependent demosaicing process for plenoptic cameras using ray optics. Our main contributions are: 1) Guidelines for future depth-dependent demosaicing approach for plenoptic cameras. 2) A framework for analyzing the demosaicing process on a focused plane is proposed based on ray-tracing.

The paper is organized as follows: we first revisit the previous demosaicing approaches for plenoptic cameras in Section 2. A detailed description of our analysis of plenoptic demosaicing is presented in Section 3. Finally, the contribution of this paper is concluded in Section 4.

2. RELATED WORK

For conventional digital cameras, image demosaicing has been widely discussed, and numerous approaches have been proposed to improve the demosaicing performance [9, 10, 11]. Generally, conventional demosaicing can be considered as an interpolation problem on the raw sensor image. These techniques are not explicitly addressing the demosaicing problems for plenoptic cameras as they neglect the unique MLA structure. Widely-used decoding pipelines still demosaic the captured lenslet images based on conventional linear demosaicing [12], resulting in undesired color aliasing artifacts [13].

Recently, David *et al.* [14] proposed a demosaicing method which tackles color fringes artifacts by using a white lenslet image. First, a white lenslet image is used for discarding the pixels that belong to different lenslets as they create crosstalk artifacts on the lenslet borders. Then gradient corrected interpolation [15] is adapted by varying the weight of neighboring pixels according to the white lenslet image. By processing each elemental image individually, this approach does not consider the contributions from other lenslets while demosaicing. As a result, the high frequency of 4D plenoptic capture is downsampled as low frequency 2D information, causing image blur on the edges.

Yu *et al.* [16] proposed to demosaic the view after rendering it on a focused plane in contrast to demosaicing the raw image on the sensor. Specifically, the radiance is first mapped to a focal plane and a frequency domain resampling is applied to ensure uniformly distributed color samples. Then the demosaicing is conducted on the refocused plane using anisotropic adaptive filtering in the frequency domain [11]. Although this approach considers the MLA structure of plenoptic cameras and greatly suppresses aliasing artifacts, it mainly focuses on the super-resolution and fails to consider the non-periodic sampling of different color channels.

3. ANALYSIS OF PLENOPTIC DEMOSAICING

In this section, we provide a theoretical analysis based on the ray-tracing technique to show that the demosaicing process for plenoptic cameras is depth-dependent and conventional demosaicing approaches cannot be applied to plenoptic image directly. For simplicity, we model each lenslet as a pinhole approximation and only the principal rays are considered.

3.1. Notation

Before proceeding with our analysis, the following notation is introduced: The radiance of a principal ray passing through a lenslet is represented as $R = (x, y, z, a, b, c)^T$ where vector $\mathbf{P} = (x, y, z)^T$ and $\mathbf{D} = (a, b, c)^T$ indicate the initial point and the direction of the ray respectively. The distance between two parallel planes Π and Π' is given by $L(\Pi, \Pi')$. For simplicity, the sensor

plane is defined as Π_s: $z = 0$. Moreover, we assume that the sensor plane, the MLA principal plane and the refocus plane are all well aligned, in other words, their positions can be denoted in the form of Π: $z = A$, where $A \in [0, +\infty)$.

3.2. Ray-tracing and phase space analysis of plenoptic capture

To initiate the ray-tracing process, N rays are generated for each pixel. Note that as the pinhole camera model is applied for MLA and the framework is linear, N can be chosen as one to save computational energy without losing the pixel sampling structure on focus planes. However, if the framework is nonlinear, more rays are required to afford a better description of the pixel sampling behavior. We consider principal rays emitting from $\mathbf{P}_s = (x_s, y_s, z_s)^T$ on the sensor that pass through the optical center $\mathbf{P}_c = (x_c, y_c, z_c)^T$ of a lenslet. $L(\Pi_c, \Pi_s) = z_c - z_s$ is the distance between the sensor plane Π_s: $z = 0$ and the MLA principal plane Π_c: $z = z_c$, the normalized direction vector D of a ray can be calculated by the following equation:

$$\mathbf{D} = \frac{\mathbf{P}_c - \mathbf{P}_s}{||\mathbf{P}_c - \mathbf{P}_s||}, \qquad (1)$$

where $||\mathbf{P}_c - \mathbf{P}_s||$ is the norm of vector $\mathbf{P}_c - \mathbf{P}_s$. Additionally, any point $\mathbf{P}$ on the ray R passing both $\mathbf{P}_s$ and $\mathbf{P}_c$ in 3D space can be explicitly specified by using weighted line representation form:

$$\mathbf{P} = (1 - t)\mathbf{P}_s + t\mathbf{P}_c, \qquad (2)$$

where the variable $t \in [0, +\infty)$ indicates the position of the point $\mathbf{P}$ on the ray, and it moves from $\mathbf{P}_s$ in the direction of $\mathbf{P}_c - \mathbf{P}_s$ as t increases.

The intersection of a ray R and an arbitrary focus plane Π: $z = A$ can be derived by calculating t from Eq. 2:

$$t = \frac{A - z_s}{z_c - z_s}. \qquad (3)$$

Thus, the intersection point can be acquired by substituting the only unknown t in Eq. 2. As mentioned in Section 3.1, we define the sensor plane as Π_s: $z = 0$, therefore Eq. 3 can be simplified as $t = \frac{A}{z_c}$. By back-projecting rays from a pixel onto the refocus plane, the sampling of pixels can be depicted, as shown in Fig. 1.

Claim 1 *Demosaicing methods developed for conventional cameras are inadequate for plenoptic cameras.*

The main challenge for conventional demosaicing schemes is to explore the relationships among neighboring pixels of the same and different color channels on the sensor to restore the color information. However, the sensor-based analysis is discrepant in the context of plenoptic cameras due to the effect of the MLA structure.

As shown in Fig. 1, rays that pass thorough the same lenslet are coded in the same color (blue or green), while the highlighted orange rays lie on the borders of two neighboring lenslet grids. As can be seen from Fig. 1, the highlighted adjacent pixels sample the focus planes with a great difference. This means that applying conventional demosaicing which disregards the MLA structure produce chromatic artifacts and erroneous interpolation results. To address this problem, some existing plenoptic demosaicing approaches apply conventional demosaicing to individual lenslets [14] as mentioned in Section 2, whereas only pixels belonging to the same lenslet are used in the demosaicing interpolation. This causes a loss resolution on the focus plane.

Another depiction of the sampling grids is a phase space representation, as shown in Fig. 2. The pixels are colored in the same manner as in Fig. 1, with q and p indicating the spatial and angular
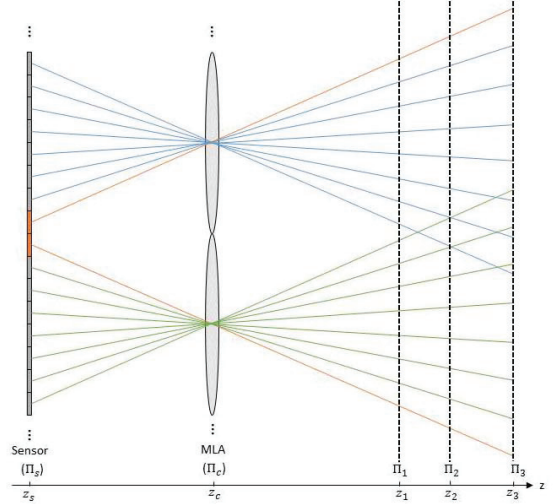

Figure 1: Sampling pattern for sensor pixels on different focus planes Π_1, Π_2 and Π_3. For simplicity, without compromising generality, only rays from the centers of the pixels in one spatial dimension are shown.

sampling range on the focus plane respectively. Note that some of the neighboring pixels on the sensor sample different spatial information and the adjacency present on the sensor no long holds on focus planes. Therefore, applying demosaicing methods that are designed for conventional cameras on plenoptic images generate erroneous color restoration and reconstruction errors [17].

Claim 2 *The demosaicing scheme for the plenoptic camera is depth-dependent (axially variant).*

Compared with conventional cameras, one of the major advantages of plenoptic capture is that it enables depth estimation. As a consequence of knowing depth, several post-processing techniques can be performed after capturing. However, by far very little work has focused on exploring the correlation between depth and demosaicing process. Here we claim that the plenoptic demosaicing is depth-dependent: We can rewrite the ray equation by substituting Eq. 3 into Eq. 2 as:

$$\mathbf{P} = \frac{z_c - A}{z_c}\mathbf{P}_s + \frac{A}{z_c}\mathbf{P}_c, \quad \frac{A}{z_c} > 1 \qquad (4)$$

thus, a ray intersects different focus planes with different lateral positions. This can be seen from Fig. 1 as the change in height of the ray, and in Fig. 2 as the skewing slope of lenslet sampling. Both vary with different focus planes. This implies that the sampling of pixels are depth-dependent and the demosaicing scheme should be depth-adaptive in order to interpolate any color channel for a spatial position.

Claim 3 *The demosaicing scheme for the plenoptic camera is laterally variant on a focus plane.*

As shown in Fig. 1, the rays are not distributed uniformly on focus planes. On the focus plane Π_1, there is an empty space which is not sampled by any pixel whereas spatial information is densely sampled elsewhere. On focus plane Π_2, two lateral positions are sampled by rays from different lenslets, as rays of different pixels reach the same position. On plane Π_3, some rays from one lenslet fall between adjacent pixels from another lenslet. This means that some regions are more densely sampled than other regions. Thus, the demosaicing method for plenoptic capture should be adapted to different lateral positions.
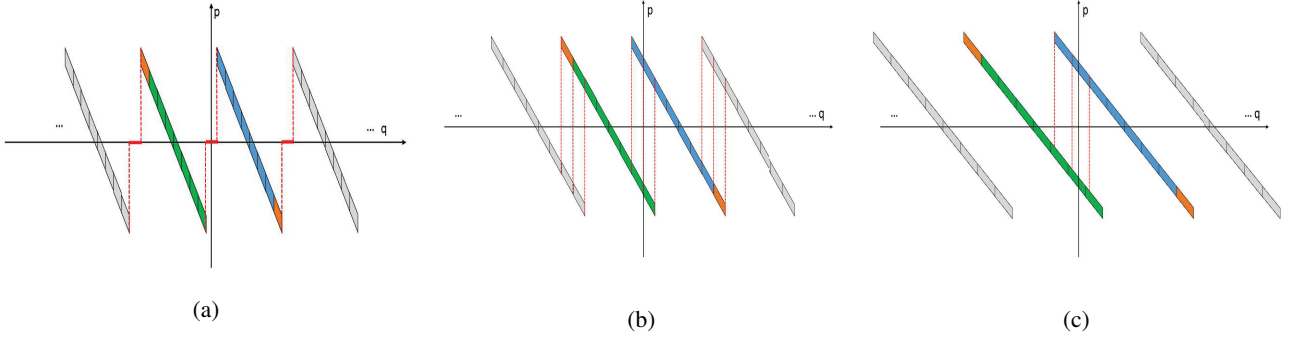
Figure 2: Phase space diagrams corresponding to different focus planes: (a) Π_1, (b) Π_2, and (c) Π_3 shown in Fig. 1, the sampling grids of four elemental images are shown.

The same conclusion can be derived from Fig. 2: Note that integrating along the p-axis on a position q gives intensity at that position on the focus plane. In Fig. 2a, not sampled spatial positions are shown as the red gap between the sampling grids of lenslets on q-axis. In Fig. 2b, the sampling range of two pixels from adjacent lenslets completely overlap on q-axis, which means that they sample the same spatial position. Another case of sampling is shown in Fig. 2c when the projection of pixels on a focus plane partially overlaps on q-axis.

Let $\mathbf{P}_i$ and $\mathbf{P}_j$ represent projections of any two pixel centers on an arbitrary focus plane, this yields:

$$d = \min\left\{||\mathbf{P}_i, \mathbf{P}_j||\right\}, \quad i \neq j \tag{5}$$

where d is the distance between $\mathbf{P}_i$ and the closest projection of any other pixel center. If knowing the pixel pitch on the sensor, denoted by K, combining with Eq. 4 we obtain:

$$K' = \left|\frac{(z_c - A)K}{z_c}\right|, \tag{6}$$

where K' is the pixel pitch when projected onto the plane Π: $z = A$. If $\mathbf{P}_i$ and $\mathbf{P}_j$ belong to pixels of the same color channel, then the demosaicing process can be described as the following: In case the sampling overlaps (fully or partially, corresponding to focus plane Π_2 and Π_3, respectively.), the resulting value can be a weighted average. In case there is no information at the actual point of interest, the value must be interpolated from adjacent pixels on the focus plane rather than adjacent ones on the sensor. The actual calculations for these two cases are very similar: In one case the weighting depends on the size of the overlapping area between known data and estimate data, in the other the weighting depends on the distance to the known data. In fact, the overlap can both be expressed as a distance, by which the two cases merge into one.

3.3. Simulation result

In order to verify our claims, a ray-tracing framework was implemented in Matlab. The focal length of MLA was set as $z_c = 46\mu m$ and the pixel pitch $K = 1.5\mu m$. For the purpose of visualization, only a 2×2 MLA structure was considered for rendering on focus planes and each elemental image was composed of a 4×4 pixel grid. All the pixels on the sensor were filtered by the bayer pattern CFA.

By projecting rays onto the focus planes, the sampling pattern of a 2×2 lenslet structure can be shown as in Fig. 3. The center of each pixel is rendered as a monochromatic asterisk of either red, green or blue. The lower insets describe the full sizes of the pixels

with their corresponding sensor coordinates in the color blocks. Note that in Fig. 3c the positions of the coordinates indicate the different centers of the pixel projection and they partly overlap on the focus plane.

When the focus plane Π: $z_1 = 70\mu m$ is placed close to the MLA Π_c: $z_c = 46\mu m$, rays of different lenslets do not intersect on the focus plane, and there is a gap between the sampling of different lenslets, as can be seen from Fig. 1. Only in this case, pixels from the same lenslet evenly and consecutively sample the near focus plane, whereas there exists unsampled area across the lenslets blocks, as shown in Fig. 3a.

It can be inferred from Eq. 6 that as the focus plane moves farther away from the MLA, the size of the pixel projection on the focus plane increases linearly. As a consequence, each pixel is sampling a larger area on the focus plane. As shown in phase diagram Fig. 2, when the projection of pixel centers from different lenslets are well aligned, such as when $K' = 3\mu m$ and $z_2 = 138\mu m$ in our setup, the sampling patterns of different microlens pixels coincide, as shown in Fig. 3b. In other cases, the pixel projection partly overlap on the focus plane, as we can see in Fig. 3c.

4. CONCLUSION AND FUTURE WORK

In this paper, we have presented the theoretical analysis for demosaicing the plenoptic capture based on ray-tracing technique. We show that demosaicing schemes for conventional cameras are not suitable for plenoptic cameras, either applying it to the individual elemental image or treat the plenoptic capture as a color image. This is due to the fact that the plenoptic camera captures the 4D radiance of the scene thanks to the MLA structure which decouple spatial and angular information, whereas a conventional camera only records planar information of the scene. Furthermore, the optimal demosaicing approach for plenoptic cameras is inherently dependent on the both depth and lateral location as a result of the plenoptic sampling. In the future, a detailed demosaicing scheme will be proposed for plenoptic capture, and the effect of wave properties of light, such as point spread function, on the plenoptic demosaicing will be investigated.

5. ACKNOWLEDGMENT
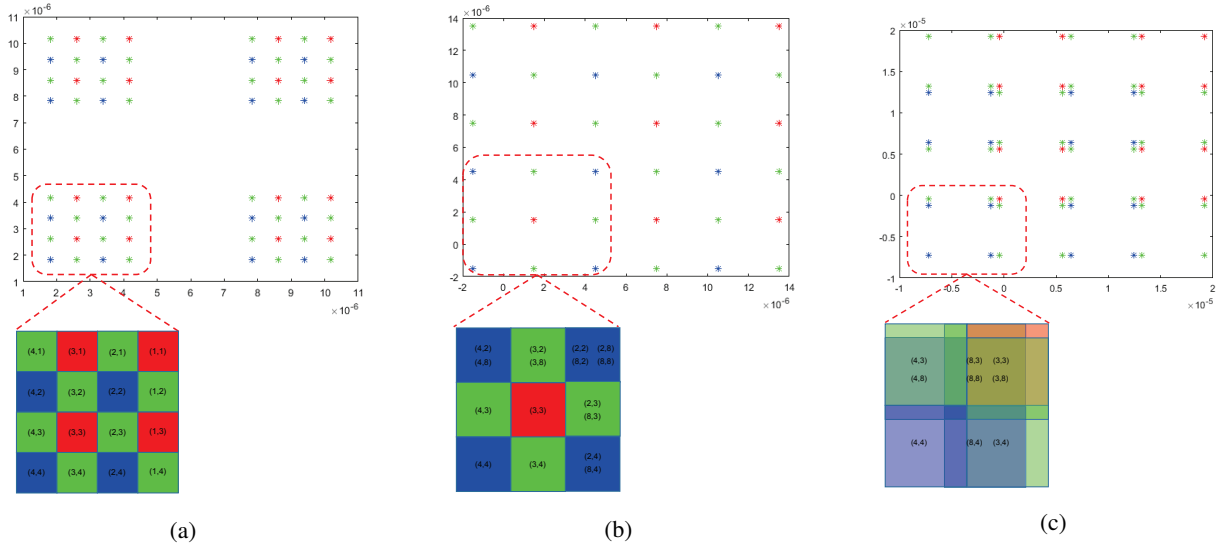
Figure 3: 2D sampling pattern of the MLA-based plenoptic camera when pixels are projected onto different focus planes: (a) $\Pi_1 : z_1 = 70\mu m$, (b) $\Pi_2 : z_2 = 138\mu m$, and (c) $\Pi_3 : z_3 = 255\mu m$, corresponding to focus planes Π_1, Π_2 and Π_3 respectively in Fig. 1.

6. REFERENCES

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

[2] Julia Navarro and Antoni Buades, "Robust and dense depth estimation for light field images," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1873–1886, 2017.

[3] Sven Wanner and Bastian Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 3, pp. 606–619, 2014.

[4] 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.

[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] Tom E Bishop, Sara Zanetti, and Paolo Favaro, "Light field superresolution," in *Computational Photography (ICCP), 2009 IEEE International Conference on*. IEEE, 2009, pp. 1–9.

[7] Todor G Georgiev and Andrew Lumsdaine, "Focused plenoptic camera and rendering," *Journal of Electronic Imaging*, vol. 19, no. 2, pp. 021106, 2010.

[8] Ron Kimmel, "Demosaicing: image reconstruction from color ccd samples," *IEEE Transactions on image processing*, vol. 8, no. 9, pp. 1221–1228, 1999.

[9] Hung-An Chang and Homer H Chen, "Stochastic color interpolation for digital cameras," *IEEE Transactions on circuits and systems for video technology*, vol. 17, no. 8, pp. 964–973, 2007.

[10] Brice Chaix De Lavarène, David Alleysson, Barthélémy Durette, and Jeanny Hérault, "Efficient demosaicing through recursive filtering," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*. IEEE, 2007, vol. 2, pp. II–189.

[11] Nai-Xiang Lian, Lanlan Chang, Yap-Peng Tan, and Vitali Zagorodnov, "Adaptive filtering for color filter array demosaicking," *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2515–2525, 2007.

[12] Donald G Dansereau, Oscar Pizarro, and Stefan B Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 1027–1034.

[13] Xiang Huang and Oliver Cossairt, "Dictionary learning based color demosaicing for plenoptic cameras," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on*. IEEE, 2014, pp. 455–460.

[14] Pierre David, Mikaël Le Pendu, and Christine Guillemot, "White lenslet image guided demosaicing for plenoptic cameras," in *MMSP 2017-IEEE 19th International Workshop on Multimedia Signal Processing*, 2017, pp. 1–6.

[15] Henrique S Malvar, Li-wei He, and Ross Cutler, "High-quality linear interpolation for demosaicing of bayer-patterned color images," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*. IEEE, 2004, vol. 3, pp. iii–485.

[16] 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.

[17] Hyunji Cho and Hoon Yoo, "Masking based demosaicking for image enhancement using plenoptic camera," *International Journal of Applied Engineering Research*, vol. 13, no. 1, pp. 273–276, 2018.