

# PlenopticCam v1.0: A Light-Field Imaging Framework

Christopher Hahne<sup>ID</sup>, Member, IEEE, and Amar Aggoun

**Abstract**—Light-field cameras play a vital role for rich 3D information retrieval in narrow range depth sensing applications. The key obstacle in composing light-fields from exposures taken by a plenoptic camera is to computationally calibrate, align and rearrange four-dimensional image data. Several attempts have been proposed to enhance the overall image quality by tailoring pipelines dedicated to particular plenoptic cameras and improving the consistency across viewpoints at the expense of high computational loads. The framework presented herein advances prior outcomes thanks to its novel micro image scale-space analysis for generic camera calibration independent of the lens specifications and its parallax-invariant, cost-effective viewpoint color equalization from optimal transport theory. Artifacts from the sensor and micro lens grid are compensated in an innovative way to enable superior quality in sub-aperture image extraction, computational refocusing and Scheimpflug rendering with sub-sampling capabilities. Benchmark comparisons using established image metrics suggest that our proposed pipeline outperforms state-of-the-art tool chains in the majority of cases. Results from a Wasserstein distance further show that our color transfer outdoes the existing transport methods. Our algorithms are released under an open-source license, offer cross-platform compatibility with few dependencies and different user interfaces. This makes the reproduction of results and experimentation with plenoptic camera technology convenient for peer researchers, developers, photographers, data scientists and others working in this field.

**Index Terms**—Plenoptic, light-field, calibration, color transfer.

## I. INTRODUCTION

THERE is a growing body of literature in the areas of experimental photography [1], [2], medical imaging [3]–[6] and machine learning [7] recognizing capabilities offered by light-fields.

### A. Background

The probably most influential light-field model in computer graphics was devised by Levoy and Hanrahan [8] who described a light-field to be a collection of ray vectors piercing through two planes stacked behind one another. The intersections of ray vectors at the two planes make up four Cartesian

Manuscript received October 22, 2020; revised March 7, 2021 and May 27, 2021; accepted June 21, 2021. Date of publication July 19, 2021; date of current version July 29, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Sérgio De Faria. (*Corresponding author: Christopher Hahne.*)

The authors are with the School of Mathematics and Computer Science, University of Wolverhampton, Wolverhampton WV1 1LY, U.K. (e-mail: c.hahne@wlv.ac.uk; a.aggoun@wlv.ac.uk).

Digital Object Identifier 10.1109/TIP.2021.3095671

coordinates, which is why light-fields are often referred to as four-dimensional (4D). In their much celebrated paper, Levoy and Hanrahan demonstrate that the two planes serve as the spatial and angular image domain providing two different light-field representations that can be transferred into each other. Such light-field transformation corresponds to changing the sequential order of the two planes.

Capturing a light-field from a monocular lens is achieved with a single sensor stacked behind an aperture grid [9] such as a Micro Lens Array (MLA). This optical setup is known as a plenoptic camera and can be thought of as accommodating an array of consistently spaced virtual cameras located at the aperture plane [10]. As opposed to light-fields from a camera array, captures from the plenoptic camera need to undergo additional processing to be represented as such [11], [12].

Until today, the landscape of software tools treating plenoptic content has been characterized by heterogeneity. One of the most mature and influential software applications was released by camera manufacturer Lytro in 2012. Lytro’s image processing pipeline remains closed-source and is not publicly maintained since the company shut down business in 2018. In the earlier days of Lytro’s lifetime, independent programmers developed binary file decoders by reverse-engineering Lytro’s file formats [13], [14]. These tools, however, are unable to perform light-field rendering functionalities such as refocusing. Several scientists published methods on how to algorithmically calibrate and decode Lytro’s plenoptic camera within Matlab [11], [12], [15], [16]. These methods are based on the given metadata while concentrating on the micro image center detection, 4D rearrangement and rectification of radial lens distortions. More recently, research has taken the direction to successfully recover physical information at light-field boundaries [17], [18]. Although these studies revealed convincing results, light-field color consistency as proposed by Matysiak *et al.* [18] requires high computational resources.

### B. Novelties

While consolidating preliminary decoding procedures, we introduce novel calibration, resampling and refocusing techniques. An overview of our contributions is listed below:

- 1) Scale space analysis of micro images is conducted early in the calibration pipeline to support plenoptic cameras with arbitrary sensor and MLA dimensions and, unlike other methods in the field, handle footage from custom-built prototypes as well as Lytro cameras.

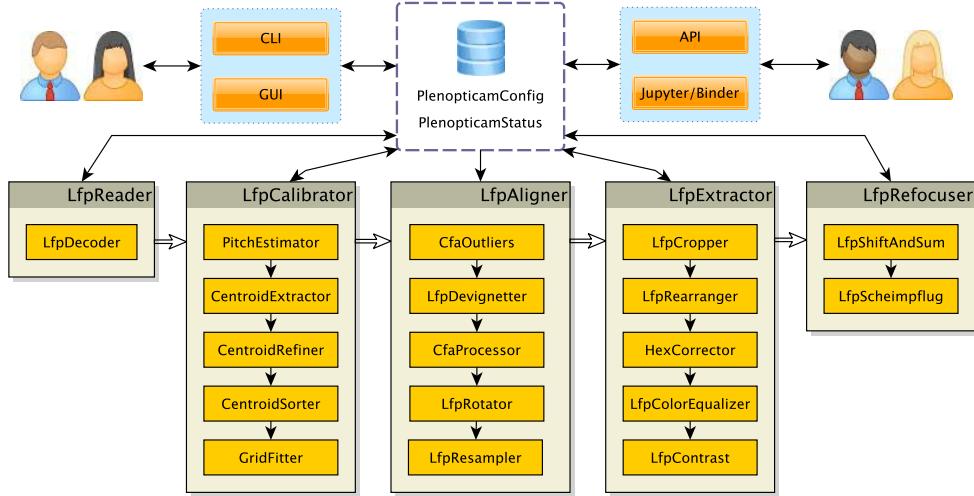


Fig. 1. **Package architecture** with gray blocks representing modules whereas orange blocks indicate essential top-level classes.

- 2) Centroid grid fitting is employed using the Levenberg-Marquardt optimization to reduce least-squares errors of detected micro lens centers globally and determine the centroid spacing and projective mapping simultaneously.
- 3) De-vignetting is based on a 4D white image least-squares fit model to suppress noise propagation as it occurs in a conventional white image division.
- 4) To counteract color variances across sub-aperture views, we propose a novel, rapid and effective illumination channel correction scheme that outperforms the previous color distribution transport techniques [19], [20].
- 5) For accurate angular sampling, we offer micro image resampling followed by a hexagonal artifact removal.
- 6) Taking advantage of the popular computational refocusing, we pave the way for its Scheimpflug equivalent.

These innovative algorithms are bundled as a tool coined PLENOPTICAM to accomplish outstanding results, which are validated in this study with standard image metrics.

### C. Scope

Equipped with generic calibration and novel rendering routines, PLENOPTICAM may build the foundation for future work on light-field image algorithms. Research goals are more easily attained as image scientists can adapt code and focus on their individual idea. In general, all kinds of plenoptic cameras are covered by PLENOPTICAM while raw buffer conversion is fully supported for the Lytro Illum. This framework yet serves as a starting point for plenoptic 2.0 and Raytrix [21] images, which share calibration and resampling requirements. Note that PLENOPTICAM can be used in conjunction with the geometry tool PLENOPTISIGN [22] to pinpoint metric object positions in a light-field captured by a plenoptic camera.

### D. Structure of the Paper

The organisation of this paper starts with an overview of the module architecture, which can be regarded as a roadmap for Section II. Subsections contained in Section II

give insightful details on novel algorithmic aspects with respect to plenoptic image calibration, sub-aperture processing and refocusing. A benchmark comparison of results rendered by PLENOPTICAM and other tool chains is carried out in Section III. Section IV draws conclusions, while reflecting on the framework's potential impact and sketching out ideas for future work.

## II. MODULE FUNCTIONALITIES

An architectural scheme of the processing pipeline is depicted in Fig. 1. Objects of a PLENOPTICAMCONFIG and PLENOPTICAMSTATUS class are thought to be singletons and shared across modules. The processing direction of input image data is from left to right (on module-level) and top to bottom (on class-level) as indicated by the arrows. The diagram in Fig. 1 can be regarded as a roadmap for this section in which the core functionality of each module is presented.

### A. LfpReader

The LFPREADER module supports standard image file types (*tiff*, *bmp*, *jpg*, *png*) and the more specific raw Lytro file type decoding (*lfp*, *lfr*, *raw*) for Bayer image composition according to the findings by Patel [13]. For raw data from a Lytro camera, this module exports an image as a *tiff* file in Bayer representation as well as a *json* file containing corresponding metadata which is used for gamma correction. Other image file types are expected to be in sRGB space for gamma handling.

### B. LfpCalibrator

The fundamental problem we aim to solve when calibrating a plenoptic camera is to register geometric properties of the 4D micro image representation. This enables a light-field to be rearranged as if captured by multiple cameras with consistent spacing [10]. PLENOPTICAM introduces a novel calibration pipeline with a sequence of steps shown in Fig. 2.

The entire calibration procedure can be sub-divided into pitch estimation, centroid extraction along with refinement,

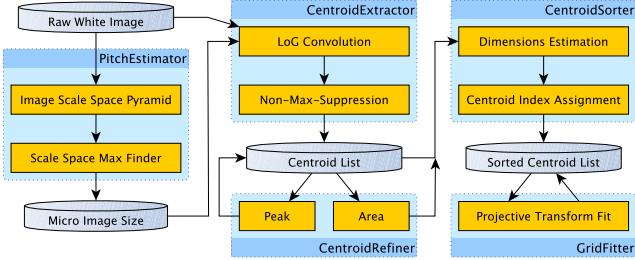


Fig. 2. Calibration pipeline.

sorting and grid fitting. To let the pipeline cover multiple types of plenoptic cameras with varying lens and sensor specifications, a blob detector is employed as a first step for analysis of the micro image size.

1) *PitchEstimator*: Based on a white calibration image  $I_w(\mathbf{x})$  where  $\mathbf{x} = [k \ l]^T \in \mathbb{N}^2$  consists of two spatial coordinates  $k, l$  across arbitrary micro images, we adapt the classical scale space theory [23] using half-octave pyramids [24] denoted by  $P(v, \mathbf{x})$  and built via

$$\forall v, P(v+1, \mathbf{x}) = D_2 \left( P(v, \mathbf{x}) * \nabla^2 G(\sigma, \mathbf{x}) \right) \quad (1)$$

with  $\{v \in \mathbb{N} \mid v < \log_2(\min(K, L))\}$  while  $K, L$  are spatial resolutions of  $I_w(\mathbf{x})$ . Downsampling is represented by  $D_2(\cdot)$  considering the half-octave requirement and  $\nabla^2 G(\sigma, \mathbf{x})$  is the Laplacian of Gaussian (LoG) convolution kernel, known as the Mexican hat, which is approximated by the Difference of Gaussians. Each pyramid level at index  $v$  is then a downsampled LoG-filtered version of the previous level where  $P(0, \mathbf{x}) = I_w(\mathbf{x}) * \nabla^2 G(\sigma, \mathbf{x})$  serves as the initial scale. Let  $v^*$  be the level exhibiting the maximum intensity across all  $v$  given by

$$v^* = \arg \max_v \left\{ \max_{\mathbf{x}} P(v, \mathbf{x}) \right\} \quad (2)$$

we infer  $\sigma^*$  as the detected blob radius at the initial scale by

$$\sigma^* = 2^{v^*/2} \sqrt{2^{\text{mod}(v^*, 2)}} \quad (3)$$

where the base-2 terms account for halved resolutions and the half-octave representation across pyramid levels, respectively. It may be an intuitive observation that a diameter  $M$  of a micro image proportionally scales with the blob radius  $\sigma^*$ . A proof for this relationship is provided in Appendix A. Section III-A demonstrates that our automatic detection of the micro image size  $M$  is an important feature for micro image registration and subsequent light-field alignment processes, which enable PLENOPTICAM to cope with a variety of different MLA and objective lens specifications.

2) *CentroidExtractor*: Initial approximation of Micro Image Centers (MICs) is a key task in plenoptic image registration and has been subject of existing research. An early method developed by Dansereau *et al.* [12] convolves a white image with a kernel of fixed size dedicated to a Lytro camera and subsequently analyzes local peaks. Cho *et al.* [11] and similarly Liang and Wang *et al.* [25] iteratively apply a morphological erosion operator to white images until a pattern of isolated micro image regions is obtained. Although image erosion facilitates rough centroid detection from intensity

maxima, the iterative nature of this approach leaves space for optimization with regard to speed and robustness. Unlike the previous detection methods, we identify MICs as local extrema in a LoG-convolved image  $I_c(\mathbf{x}) = I_w(\mathbf{x}) * \nabla^2 G(\sigma^*, \mathbf{x})$  by using Non-Maximum-Suppression (NMS) in the form of the  $3 \times 3$  neighborhood scan. Here the kernel  $\nabla^2 G(\sigma^*, \mathbf{x})$  scales with  $M$  and helps carve out micro image peaks as it is responsive to the detected feature size. We employ Pham's NMS method [26] and denote an MIC by  $\mathbf{c}_n = [k_n, l_n]^T \in \mathbb{Z}^2$ .

3) *CentroidRefiner*: It is important to note that the preliminary CENTROIDEEXTRACTOR yields centroids located at integer sensor coordinates. Several studies have shown that integer coordinate precision induces artifacts when decomposing a raw plenoptic image to the sub-aperture image representation of a light-field [11], [12], [25]. For accurate light-field decomposition, we refine centroids with sub-pixel precision by taking pixel intensities into account that belong to the same micro image region  $\mathcal{R}_n$ , which is typically in a range of  $M \times M$  size. To obtain sub-pixel precise centroids  $\bar{\mathbf{c}}_n = [\bar{k}_n \ \bar{l}_n]^T \in \mathbb{R}^2$ , we perform coordinate refinement by

$$\bar{k}_n = \sum_{(k,l) \in \mathcal{R}_n} \frac{I_c(k, l) \cdot k}{\sum I_c(k, l)} \text{ and } \bar{l}_n = \sum_{(k,l) \in \mathcal{R}_n} \frac{I_c(k, l) \cdot l}{\sum I_c(k, l)} \quad (4)$$

where  $I_c(k, l)$  represents a LoG-convoluted white image. As an alternative, we compute area centroids from a binary micro image after thresholding with the 75th percentile of  $I_c(k, l)$  which then simplifies Eq. (4) to

$$\bar{k}_n = \frac{1}{|\mathcal{R}_n|} \sum_{(k,l) \in \mathcal{R}_n} k \text{ and } \bar{l}_n = \frac{1}{|\mathcal{R}_n|} \sum_{(k,l) \in \mathcal{R}_n} l \quad (5)$$

where  $|\cdot|$  denotes the cardinality providing the total number of elements within a micro image region  $\mathcal{R}_n$  above the threshold value. By default, PLENOPTICAM uses the latter approach given in Eq. (5) as it proves to be the more generic solution for white images suffering from noise or saturation. Figures 5 and 7 show examples of detected centroids using this method.

4) *CentroidSorter*: At this point, only little is known about the dimensions and geometric micro lens arrangement. To enable generic calibration for different camera models with custom MLAs and arbitrary lens numbers, it is mandatory to examine such fundamental properties. Let  $\mathbf{C} = \{\bar{\mathbf{c}}_n \mid n \in \mathbb{N}\}$  be a finite and unordered set of centroids  $\bar{\mathbf{c}}_n = [\bar{k}_n \ \bar{l}_n]^T$ . For a first centroid spacing approximation, we assume the ratio of sensor dimensions  $K$  and  $L$  matches the aspect ratio of the MLA giving  $H = \sqrt{|\mathbf{C}| \times L/K}$  for the horizontal micro lens resolution with  $|\mathbf{C}|$  as the total number of centroids. The centroid spacing is estimated via  $L/H$  and used to form a set of neighbours  $\mathbf{S}$  from an arbitrary centroid  $\bar{\mathbf{c}}_r \in \mathbf{C}$  via

$$\mathbf{S} = \{\bar{\mathbf{c}}_n - \bar{\mathbf{c}}_r \mid \bar{\mathbf{c}}_n \in \mathbf{C} \wedge L/2H < \|\bar{\mathbf{c}}_n - \bar{\mathbf{c}}_r\|_2 < 3L/2H\} \quad (6)$$

where  $\|\cdot\|_2$  denotes the  $\ell^2$  norm. The MLA packing geometry  $\mathcal{P}$  is determined by analyzing angles  $\alpha_n$  of  $\bar{\mathbf{c}}_n \in \mathbf{S}$  as follows

$$\mathcal{P} := \begin{cases} \text{hexagonal,} & \text{if } \forall \bar{\mathbf{c}}_n \in \mathbf{S}, \lfloor \alpha_n 12/\pi \rfloor \in \{1, 3\} \\ \text{rectangular,} & \text{if } \forall \bar{\mathbf{c}}_n \in \mathbf{S}, \lfloor \alpha_n 12/\pi \rfloor \in \{0, 3\} \end{cases} \quad (7)$$

where  $\lfloor \cdot \rfloor$  is the nearest integer operation and  $\alpha_n$  is given by

$$\alpha_n = \arccos \left( \frac{(\mathbf{1}^\top \bar{\mathbf{c}}_n)}{\sqrt{(\mathbf{1}^\top \mathbf{1})(\bar{\mathbf{c}}_n^\top \bar{\mathbf{c}}_n)}} \right) \quad (8)$$

while  $\mathbf{1} = [1 \ 1]^\top \in \mathbb{R}^2$  acts as a reference vector at  $45^\circ$ .

Rearranging plenoptic micro images to a sub-aperture light-field requires centroids to be indexed in 2D since their relative positions act as spatial coordinates in the sub-aperture image domain. However, the order of centroids within the array remains ambiguous. Thus, we seek a procedure that assigns indices  $j \in \{1, 2, \dots, J\}$  and  $h \in \{1, 2, \dots, H\}$  to each micro image centroid  $\bar{\mathbf{c}}_{j,h}$ . The proposed sort procedure begins with the search for the most upper left centroid  $\bar{\mathbf{c}}_{1,1}$  with  $j = 1, h = 1$  which is found by the minimum Euclidean distance to the image origin as given by

$$\bar{\mathbf{c}}_{1,1} = \bar{\mathbf{c}}_{n^*}, \quad \text{where } n^* \in \arg \min_{1 \leq n \leq |\mathbf{C}|} (\|\bar{\mathbf{c}}_n\|_2) \quad (9)$$

On the basis of an indexed centroid  $\bar{\mathbf{c}}_{j,h}$ , we search for its spatial neighbor by iterating through  $\mathbf{C}$  until finding a candidate  $\bar{\mathbf{c}}_n = [\bar{k}_n \ \bar{l}_n]^\top$  as seen in

$$\bar{\mathbf{c}}_{j+1,h} = \{\bar{\mathbf{c}}_n \mid \bar{\mathbf{c}}_n \in \mathbf{C} \wedge \varphi(\bar{\mathbf{c}}_{j,h}, \bar{\mathbf{c}}_n)\} \quad (10)$$

that satisfies a boundary condition  $\varphi(\cdot) \in \{0, 1\}$  given by

$$\begin{aligned} \varphi(\bar{\mathbf{c}}_a, \bar{\mathbf{c}}_b) &= \bar{k}_a + \Upsilon_1 < \bar{k}_b < \bar{k}_a + \Upsilon_2 \\ &\wedge \bar{l}_a + \Upsilon_3 < \bar{l}_b < \bar{l}_a + \Upsilon_4 \end{aligned} \quad (11)$$

where  $\Upsilon = \mathbf{e}M/2$  with  $\mathbf{e} = [e_1 \ e_2 \ e_3 \ e_4]^\top$  is a boundary scale vector considering earlier identified MLA properties. Note that horizontal and vertical coordinates in Eq. (11) may be swapped to switch between search directions. An exemplary result showing indexed centroids is depicted in Fig. 5.

5) *GridFitter*: At this stage of the calibration pipeline, centroids rely on intensity distributions, which may be affected by a broad range of irregularities arising from the MLA-sensor compound. Dansereau *et al.* [12] compress centroid information by forming a consistently spaced grid based on provided metadata. Usage of Delaunay triangulation as proposed by Cho *et al.* [11] suits hexagonal arrangements, but lacks to compensate for grid inconsistencies. Instead, we elaborate on a least-squares (LSQ) regression as potentially intended in a patent filed by Liang and Wang [25].

Based on previously determined micro lens numbers ( $J, H$ ) and geometric packing  $\mathcal{P}$ , we tailor a grid model function  $G(\cdot)$  to produce an ordered array of points  $\tilde{\mathbf{g}}_{j,h} = G(J, H; \mathcal{P})$  where  $\tilde{\mathbf{g}}_{j,h} = [\tilde{k}_{j,h} \ \tilde{l}_{j,h} \ \tilde{z}_{j,h}]^\top \in \mathbb{R}^3$  consists of consistently spaced and normalized spatial centroid coordinates  $\tilde{k}_{j,h}$ ,  $\tilde{l}_{j,h}$  and  $\tilde{z}_{j,h}$  that may be homogenized, i.e.,  $\tilde{z}_{j,h} = 1$ . The grid generation is followed by a projective transformation using  $\mathbf{P} \in \mathbb{R}^{3 \times 3}$  that yields a 3-vector centroid  $\tilde{\mathbf{c}}'_{j,h} = \mathbf{P}\tilde{\mathbf{g}}_{j,h}$ , which in matrix notation is written as

$$\begin{bmatrix} \tilde{k}'_{j,h} \\ \tilde{l}'_{j,h} \\ \tilde{z}'_{j,h} \end{bmatrix} = \begin{bmatrix} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \\ p_7 & p_8 & 1 \end{bmatrix} \begin{bmatrix} \tilde{k}_{j,h} \\ \tilde{l}_{j,h} \\ \tilde{z}_{j,h} \end{bmatrix} \quad (12)$$

covering spatial center offsets, grid scales and tilts about three axes [27]. A centroid estimate  $\hat{\mathbf{c}}_{j,h} = [\hat{k}_{j,h} \ \hat{l}_{j,h} \ 1]^\top \in \mathbb{R}^3$  is then obtained by  $\hat{\mathbf{c}}_{j,h} = \mathbf{P}\tilde{\mathbf{g}}_{j,h}/\tilde{z}'_{j,h}$ .

To determine an optimal  $\mathbf{P}^*$ , we employ a distance metric in  $\mathbf{F}_{j,h}$  comparing all measured  $\bar{\mathbf{c}}_{j,h}$  and generated  $\hat{\mathbf{c}}_{j,h}$  by

$$\mathbf{F}_{j,h} = \|\bar{\mathbf{c}}_{j,h} - \hat{\mathbf{c}}_{j,h}\|_2 + \beta R(\bar{\mathbf{c}}_{j,h}, \hat{\mathbf{c}}_{j,h}, M), \quad \forall j, h \quad (13)$$

with a regularization term  $R(\cdot)$  and adjustable weight  $\beta$ .

Here, the regularizer penalizes false centroid shifts caused by asymmetric vignetting at off-center micro images by

$$R(\bar{\mathbf{c}}_{j,h}, \hat{\mathbf{c}}_{j,h}, M) = \begin{cases} 0, & \text{if } \bar{\mathbf{d}}_{j,h} + M/\hat{M} < 0 \\ \sum_{(\bar{k}, \bar{l})} \bar{\mathbf{d}}_{j,h}, & \text{otherwise} \end{cases} \quad (14)$$

with  $\bar{\mathbf{d}}_{j,h} = |\bar{\mathbf{c}}_{j,h} - (p_3, p_6)| - |\hat{\mathbf{c}}_{j,h} - (p_3, p_6)|$  as an auxiliary distance measure and  $\hat{M} \approx 20$  as a micro image size divider.

Let  $\mathbf{P} \in \mathbb{R}^{3 \times 3} \rightarrow \mathbf{p} \in \mathbb{R}^9$  be reshaped to a parameter vector  $\mathbf{p} = [p_1 \ p_2 \ p_3 \ p_4 \ p_5 \ p_6 \ p_7 \ p_8 \ 1]^\top$  and similarly  $\mathbf{F}_{j,h} \in \mathbb{R}^{J \times H} \rightarrow \mathbf{f} \in \mathbb{R}^{|\mathbf{C}|}$  be flattened to a vector-valued cost function  $\mathbf{f} = [f_1 \ f_2 \ \dots \ f_{|\mathbf{C}|}]^\top$ , the objective function is

$$\arg \min_{\mathbf{p}} \sum_{n=1}^{|\mathbf{C}|} f_n \quad (15)$$

for which we employ the Levenberg-Marquardt (LM) step by

$$\mathbf{p}_{k+1} = \mathbf{p}_k - (\mathbf{J}^\top \mathbf{J} + \mu \mathbf{D}^\top \mathbf{D})^{-1} \mathbf{J}^\top \mathbf{f} \quad (16)$$

where  $\mathbf{J}$  is the Jacobian,  $\mathbf{J}^\top \mathbf{J}$  approximates the Hessian with  $\top$  as the matrix transpose and  $\mathbf{J}^\top \mathbf{f}$  acts as the gradient [28]. Here, a diagonal matrix  $\mathbf{D}$  with an adaptive damping term  $\mu$  allows for fast LM convergence. As opposed to an analytically derived Jacobian  $\mathbf{J}$ , a multi-variate numerical approximate  $\tilde{\mathbf{J}}(\mathbf{p}) \approx \mathbf{J}$  is employed. The iterative update procedure proceeds until a convergence condition is met. Successful estimates  $\mathbf{P}^*$  are fed into the projective matrix for a final centroid assignment  $\hat{\mathbf{c}}_{j,h}^* = \mathbf{P}^* \tilde{\mathbf{g}}_{j,h}/\tilde{z}'_{j,h}$ . Enhancements of the herein proposed LSQ grid regression are examined in Section III-A.

### C. LfpAligner

1) *CfaOutliers*: Often, hot and dead pixels arise from electrical response variations in the sensor hardware and become noticeable as intensity outliers. In contrast to other pipelines [12], [18], we detect and rectify outliers prior to demosaicing channels from the Color Filter Array (CFA). The reasoning behind our decision is that false intensities are propagated to adjacent pixels during demosaicing, which may pass unnoticed by a detection at a later processing stage. To identify outliers, we regard each of the four Bayer channels as a two-dimensional (2D) grey scale image  $I_B(\mathbf{x})$  and analyze the difference from its Median-filtered version given by

$$I_R(\mathbf{x}) = I_B(\mathbf{x}) - \mathcal{M}(I_B(\mathbf{x})) \quad (17)$$

where  $I_R(\mathbf{x})$  is the reference image used for further analysis and  $\mathcal{M}(\cdot)$  denotes the Median filter operator. From  $I_R(\mathbf{x})$ , we obtain the arithmetic mean  $\bar{I}_R$  and the standard deviation

$\sigma_R$  as local statistical measures from a sliding window of  $2n+1$  size to replace potential outliers  $I_B(\mathbf{x})$  as follows

$$I_B(\mathbf{x}) = \begin{cases} \mathcal{M}(I_B(\mathbf{x})), & \text{if } I_B(\mathbf{x}) > \bar{I}_R + 4\sigma_R \\ I_B(\mathbf{x}), & \text{otherwise} \end{cases} \quad (18)$$

where the if clause statement helps detect intensity outliers. Once a condition is fulfilled, nearby intensities in the range of  $2n+1$  are used to replace an outlier. The selection is further constrained by only accepting a small number  $n$  of candidates in a  $n^2 \times n^2$  window. Without this constraint, many pixels of a saturated image area would be falsely detected as outliers.

2) *LfpDevignetter*: In optical imaging, vignetting occurs at non-paraxial image areas as a result of either mechanical blocking of light or illumination fall-off from the cosine-fourth law, which arise from the Lambertian reflectance, pupil size reduction and the distance-dependant inverse square law [29]. Plenoptic images suffer from vignetting in similar ways whereas the appearance is given in 4D representation. So far, micro image vignetting is treated by a pixel-wise division with a normalized white image [12], [18]. PLENOPTICAM uses this method by default, as it works well for white images not severely suffering from noise. To combat potential noise, LFPDEVIGNETTER offers alternative 4D de-vignetting based on LSQ fitting inspired by classical 2D image processing that has not yet been applied to plenoptic images. We adapt the classical procedure by iterating through each micro image and dividing it with normalized LSQ fit values to prevent noise propagation during the light-field de-vignetting.

The intensity surface of a white micro image is given as  $I_m(u, v)$  and approximated by a multivariate polynomial regression. For 2<sup>nd</sup> order polynomials, this fit function writes

$$I_m(u, v) = w_1 + w_2 u + w_3 v + \dots + w_7 u^2 v^2 \quad (19)$$

with  $\mathbf{w} = [w_1 \ w_2 \ \dots \ w_7]^T$  as the regression coefficients. Provided the 2D micro image coordinate indices  $u$  and  $v$ , this is translated to matrix form by

$$\mathbf{A} = \begin{bmatrix} 1 & u_1 & v_1 & \dots & u_1^2 & v_1^2 \\ 1 & u_2 & v_2 & \dots & u_2^2 & v_2^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & u_M & v_M & \dots & u_M^2 & v_M^2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} I_m(u_1, v_1) \\ I_m(u_1, v_2) \\ \vdots \\ I_m(u_M, v_M) \end{bmatrix} \quad (20)$$

with  $\mathbf{A}$  as the Vandermonde matrix and  $\mathbf{b}$  containing micro image intensities. With  $\mathbf{A}$  generally being non-square, the equation system is solved via the pseudo-inverse  $+$  given by

$$\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \quad (21)$$

so that we obtain fit values  $\mathbf{w}$  for each micro image by

$$\mathbf{w} = \mathbf{A}^+ \mathbf{b} \quad (22)$$

After estimation of weight coefficients, we divide each micro image by its fitted white image counterpart. The effectiveness of the above method is demonstrated in Section III-C.

3) *CfaProcessor*: The CFAPROCESSOR class is dedicated to raw sensor images taking care of debayering using Menon's algorithm [30], white balancing and color correction. Previous findings made with regards to highlight processing [18] were adopted as they yield enhanced sub-aperture image quality.

4) *LfpRotator*: At the assembling stage of a plenoptic camera, a micro lens grid is ideally placed so that its tilt angles are in line with that of a sensor array. In the real world, however, this grid may likely be displaced with respect to the pixel grid. To counteract rotations about the  $z$  axis, our approach exploits the fact that images exposing aberrations tend to be aberration-free along their central axes. This suggests that a centroid row close to the image center forms a line, which may be a reliable indicator for the MLA rotation angle. Such a central row is obtained by  $\hat{\mathbf{c}}_{j,\delta}^* = [\bar{k}_{j,\delta}^* \ \bar{l}_{j,\delta}^*]^T \in \mathbb{R}^2$  with  $\delta = \lfloor (H-1)/2 \rfloor$  and  $H$  as the total number of micro lenses in the vertical direction. Applying LSQ regression on  $\hat{\mathbf{c}}_{j,\delta}^*$  similar to Eqs. (13) and (16) yields an angle  $\theta_z$ . Centroid and image rotation is accomplished by successively multiplying coordinates with a rotation matrix  $\mathbf{R}_z$  and a translation matrix  $\mathbf{T}$  giving a rotationally aligned centroid set  $\hat{\mathbf{c}}'_{j,h} = \mathbf{T}^{-1} \mathbf{R}_z \mathbf{T} \hat{\mathbf{c}}_{j,h}^*$  for further processing. The LFPROTATOR is left optional as rotational alignments are covered in the more general resampling stage.

5) *LfpResampler*: Micro image centroids  $\hat{\mathbf{c}}_{j,h}^* \in \mathbb{R}^2$  are likely to be found at sub-pixel positions, as Section II-B reveals. Taking pixel intensities from nearest integer coordinates leads to errors and thus image artifacts [11]. It is mandatory to preserve fractional digits of centroids to accurately decompose sub-aperture images. Therefore, we provide two competitive alignment schemes, which retain geometric properties via

- 1) aligning the entire light-field by a single transformation
- 2) resampling each micro image  $I_m(u, v)$  individually

For global alignment, the projective matrix  $\mathbf{P}^* \in \mathbb{R}^{3 \times 3}$  from the GRIDFITTER in Section II-B.5 is used along with another projection matrix  $\mathbf{P}_g$  representing a desired target grid  $\tilde{\mathbf{g}}_{j,h}$  at consistently spaced pixel coordinates. The ideal global transfer matrix is then  $\mathbf{P}_t = \mathbf{P}_g (\mathbf{P}^*)^{-1}$  having 8 degrees of freedom. Similar to [12], our transfer matrix  $\mathbf{P}_t$  dictates a global light-field transformation interpolating all micro images such that their centroids exhibit consistent spacing and coincide with actual pixel centers afterwards. Note that this procedure accounts for rotational MLA deviations and facilitates decomposition into sub-aperture images at a later stage.

As an alternative, resampling is conducted locally by 2D interpolation of each micro image so that its central pixel and detected centroid match after spatial shifting. Figure 3(a) depicts local resampling showing a centroid  $\hat{\mathbf{c}}_{j,h}^*$  in a micro image with weighting coefficients  $\gamma$  from surrounding pixels.

Following the detection of a hexagonal MLA using Eq. (7), LFPRESAMPLER takes care of the conversion to a rectangular grid, which exploits the fact that three hexagonally arranged micro image centroids span an equilateral triangle with equal distance to its barycenter as seen in Fig. 3(b). Considering this geometry, a rectangular micro image grid can be obtained by averaging every set of three adjacent micro images. Figure 3(b)

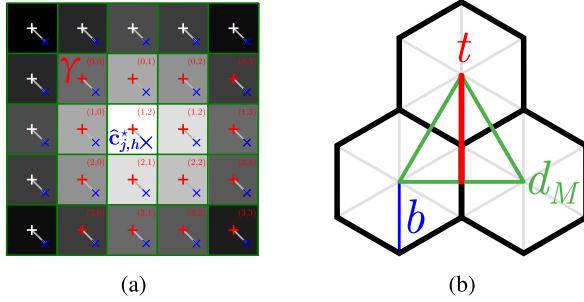


Fig. 3. **Interpolation schemes** with (a) centroid  $\hat{c}_{j,h}^*$  in a micro image and (b) hexagonal grid geometry with  $t$  as the height and  $d_M$  as the side length of an equilateral triangle.

illustrates this with  $d_M = b\sqrt{3}$  as the spacing of two centroids and the side length of a triangle whereas  $t = 3b/2$  is its height in a rectangular grid. One may note that  $d_M/t = \sqrt{3}/3/2$  becomes  $2/\sqrt{3}$  after rearranging. From this, it follows that the sampling density along  $t$  is  $2/\sqrt{3}$  times higher in relation to  $d_M$ . Therefore, upsampling the less dense spatial dimension by  $2/\sqrt{3}$  will achieve consistency in the sampling density.

In local resampling, rectangular grid conversion is accomplished by de-interleaving and elongating the dimension orthogonal to  $t$ , which corresponds to the horizontal direction in Lytro cameras. This shift and stretch alignment breaks down to translating every other coordinate vector by  $d_M/2$  and simultaneously upsampling the micro image number by  $2/\sqrt{3}$ . The light-field array is reshaped prior to the spatial interpolation as consecutive pixels are angular neighbors on the sensor.

#### D. LfpExtractor

1) *LfpRearranger*: Let  $E_{fs}[s_j, u_{c+i}]$  be a spatio-angular 2D slice of an aligned 4D micro image array, the rearrangement to a sub-aperture slice  $E_i[s_j]$  at view position  $i$  reads

$$E_i[s_j] = E_{fs}[s_j, u_{c+i}] \quad (23)$$

where micro image pixels at  $u_{c+i}$  are consecutively collected and relocated in a new vector  $E_i[s_j]$ . The preliminary resampling alignment enables each centroid to be represented by  $c = (M-1)/2$  with an odd micro image pixel diameter  $M$  so that  $\text{mod}(M, 2) = 1$ . A change in index  $i \in \{-c, \dots, c\}$  controls the relative view location in the light-field whereas index  $j \in \{1, 2, \dots, J\}$  iterates through micro lenses as the spatial domain. Note that Eq. (23) refers to one direction, which is equally applied orthogonal to it for complete rendering.

2) *HexCorrector*: As described in the LFPRESAMPLER, the elongation scheme in local resampling implies shifting every other row by half the centroid spacing to form a consistent rectangular lattice. However, this technique causes a shortcoming becoming visible as zipper-like artifacts along straight object edges in sub-aperture images (e.g., see Fig. 11).

Peers addressed this issue via barycentric interpolation [11], posterior demosaicing [31] and depth-guided resampling [32]. We tackle this artifact after the fact by the identification of affected pixels. For the detection, we compose two auxiliary images  $E_i[\hat{s}_j]$  and  $E_i[\check{s}_j]$  by de-interlacing, i.e. taking every other row, of a sub-aperture image  $E_i[s_j]$  where  $E_i[\hat{s}_j]$  are

the unshifted pixel vectors and  $E_i[\check{s}_j]$  the shifted counterparts. A pixel-wise subtraction yields local variances  $\tilde{V}_i[\check{s}_j]$  given by

$$\tilde{V}_i[\check{s}_j] = E_i[\check{s}_j] - E_i[\hat{s}_j] \quad (24)$$

which contain strong responses at edges parallel to the shift and stretch direction. To neglect real object edges, we further subtract the magnitude of the partial derivative  $\partial_{\hat{s}} E_i[\hat{s}_j]$  by

$$\bar{V}_i[\check{s}_j] = |\tilde{V}_i[\check{s}_j]| - |\partial_{\hat{s}} E_i[\hat{s}_j]| \quad (25)$$

where it is assumed that  $\bar{V}_i[\check{s}_j]$  exhibits peaks for potential candidates. Noisy responses are eliminated via threshold  $\tau$  by

$$V_i[\check{s}_j] = \begin{cases} 1, & \text{if } \bar{V}_i[\check{s}_j] > \tau \\ 0, & \text{otherwise} \end{cases} \quad (26)$$

To further exclude false positives along the shift direction, we reject candidates in  $V_i[\check{s}_j]$  not being part of a consecutive sequence of minimum length 4. In doing so, treated areas have a sufficient size to be visually recognized. The remainders in  $V_i[\check{s}_j]$  are then used to make substitutions in  $E_i[s_j]$  by

$$E_i[s_j] = \begin{cases} \frac{1}{|\mathcal{R}|} \sum_{t_h \in \mathcal{R}} E_i[t_h], & \text{if } V_i[\check{s}_j] = 1 \\ E_i[s_j], & \text{otherwise} \end{cases} \quad (27)$$

for each  $j \in \{1, 2, \dots, J\}$  while  $t_h$  is the orthogonal counterpart of  $s_j$  in a 2D region  $\mathcal{R}$ . The results are seen in Fig. 11.

3) *LfpColorEqualizer*: Lens components generally expose a gradual intensity decline toward image edges caused by the cosine-fourth law [29]. With a micro image fully covered by the sensor, this illumination fall-off is spread across a relatively small group of Bayer pattern pixels merged during demosaicing. Thus, visible intensity variances in off-axis sub-aperture views arise from micro image vignetting that we aim to rectify. Here, the goal is to propagate trusted intensities from paraxial image areas that expose no severe image aberrations to sub-aperture images located at the edge of a light-field. It can thus be used in addition to de-vignetting by utilizing the information redundancy within the light-field. A recent study employed Gaussian Mixture Models (GMMs) in conjunction with disparity correspondences computed at a preceding stage to reduce color inconsistencies [18]. However, the authors note that their procedure requires a high computational load (240 minutes per light-field) due to the many steps involved.

To overcome this limitation, we employ linear mappings based on probability distributions. Histogram Matching (HM) may be a starting point since it is invariant of the texture, preserves parallax information while requiring relatively little computational effort as opposed to iterative methods [18]. Besides using HM, we extend a recent advancement in image color transfer based on the Monge-Kantorovich-Linearization (MKL). Pitié and Kokaram [19] introduced MKL to the field of image processing to facilitate automatic color grading in media production. As of now, MKL has not been applied in the context of light-field color transfer and appears to not have been combined with channel-wise HM. Due to MKL using Multi-Variate Gaussian Distributions (MVGDs) in combination with HM, we expect our novel HM-MKL-HM compound to outperform a stand-alone HM and Pitié's pure MKL in terms of accuracy while largely

reducing the computational complexity imposed by Matysiak's method [18]. Mathematical details are presented hereafter.

Let  $\mathbf{r} \in \mathbb{R}^{JH}$  be single color channel intensities of an image  $E_i[s_j]$ , its Probability Density Function (PDF) is given by

$$f(k, \mathbf{r}) = \frac{\eta(k, \mathbf{r})}{JH}, \quad \forall k \in \{1, 2, \dots, \mathcal{L}\} \quad (28)$$

where  $\eta(k, \cdot)$  yields the number of pixels with intensity level  $k$  while  $\mathcal{L}$  is the maximum level [33]. The histogram is normalized by  $JH$  as the total pixel count of the image. From this we compute  $F(k, \mathbf{r})$  as the Cumulative Density Function (CDF), which is obtained using an auxiliary index  $\epsilon$  via

$$F(k, \mathbf{r}) = \sum_{\epsilon=1}^k \frac{\eta(\epsilon, \mathbf{r})}{JH}, \quad \forall k \in \{1, 2, \dots, \mathcal{L}\} \quad (29)$$

To match a source  $F(k, \mathbf{r})$  with a target  $G(k, \mathbf{z}) \stackrel{\text{def}}{=} F(k, \mathbf{z})$ , we perform a mapping with  $T : F(k, \mathbf{r}) \rightarrow G(k, \mathbf{z})$  that yields

$$\mathbf{r} \leftarrow T(F(k, \mathbf{r})) \quad (30)$$

where each  $F(k, \mathbf{r})$  gets assigned a new value  $G(k, \mathbf{z})$  from the probability map  $T$  implemented as a discrete lookup table.

While a channel-wise HM is effective, it fails to transfer colors at a level satisfying our visual perception. It is thus our goal to determine an optimal PDF transport. Due to the stochastic nature of intensity distributions  $(r, g, b)$ , we regard each source  $\mathbf{R} = [\mathbf{r}^{(r)} \mathbf{r}^{(g)} \mathbf{r}^{(b)}]^\top$  and target  $\mathbf{Z} = [\mathbf{z}^{(r)} \mathbf{z}^{(g)} \mathbf{z}^{(b)}]^\top$  as a correlated channel MVGD  $\mathcal{N}(\cdot)$  given by

$$\mathcal{N}(\mathbf{R}; \boldsymbol{\mu}_r, \Sigma_r) = \frac{\exp(-\frac{1}{2}(\mathbf{R} - \boldsymbol{\mu}_r)^\top \Sigma_r^{-1} (\mathbf{R} - \boldsymbol{\mu}_r))}{\sqrt{(2\pi)^{\text{rank}(\Sigma_r)} |\Sigma_r|}} \quad (31)$$

where  $\Sigma_r \in \mathbb{R}^{3 \times 3}$  and  $\Sigma_z \in \mathbb{R}^{3 \times 3}$  denote covariance matrices with  $\boldsymbol{\mu}_r \in \mathbb{R}^{3 \times 1}$  and  $\boldsymbol{\mu}_z \in \mathbb{R}^{3 \times 1}$  as the mean vectors of  $\mathbf{R} \in \mathbb{R}^{3 \times JH}$  and  $\mathbf{Z} \in \mathbb{R}^{3 \times N}$ . A desired transfer  $\hat{t}(\mathbf{R})$  requires MVGDs to be  $\mathcal{N}(\mathbf{Z}; \boldsymbol{\mu}_z, \Sigma_z) \propto \mathcal{N}(\hat{t}(\mathbf{R}); \boldsymbol{\mu}_z, \Sigma_z)$  so that a result  $\hat{t}(\mathbf{R})$  is substituted for a  $\mathbf{Z}$  when setting

$$(\mathbf{Z} - \boldsymbol{\mu}_z)^\top \Sigma_z^{-1} (\hat{t}(\mathbf{R}) - \boldsymbol{\mu}_z) = (\mathbf{R} - \boldsymbol{\mu}_r)^\top \Sigma_r^{-1} (\mathbf{R} - \boldsymbol{\mu}_r) \quad (32)$$

after dropping the constant leading terms. Using the pseudo-inverse  $+$  as in Eq. (21), we arrange the above equation to

$$\hat{t}(\mathbf{R}) - \boldsymbol{\mu}_z = \left( (\mathbf{Z} - \boldsymbol{\mu}_z)^\top \Sigma_z^{-1} \right)^+ (\mathbf{R} - \boldsymbol{\mu}_r)^\top \Sigma_r^{-1} (\mathbf{R} - \boldsymbol{\mu}_r) \quad (33)$$

As we seek a compact transfer matrix, we define  $\mathbf{M}$  to be

$$\mathbf{M} = \left( (\mathbf{Z} - \boldsymbol{\mu}_z)^\top \Sigma_z^{-1} \right)^+ (\mathbf{R} - \boldsymbol{\mu}_r)^\top \Sigma_r^{-1} \quad (34)$$

so that after substitution and rearranging, Eq. (33) becomes

$$\hat{t}(\mathbf{R}) = \mathbf{M}(\mathbf{R} - \boldsymbol{\mu}_r) + \boldsymbol{\mu}_z \quad (35)$$

for the forward MVGD transfer. The determination of  $\mathbf{M}$  is key for an optimal transport. Pitié and Kokaram employed MKL with  $\mathbf{M}^\top \Sigma_z^{-1} \mathbf{M} = \Sigma_r^{-1}$  which is given by

$$\mathbf{M} = \Sigma_r^{-1/2} \left( \Sigma_r^{1/2} \Sigma_z \Sigma_r^{1/2} \right)^{1/2} \Sigma_r^{-1/2} \quad (36)$$

One may note that the authors elaborated on a variety of solutions for  $\mathbf{M}$  (e.g., Cholesky factorization) among which

MKL proved to be the most successful in terms of accuracy [34]. We complement this concept by providing an analytical solution in Eq. (34) and combining it with HM in a sequential order (e.g., HM-MVGD-HM). Taking the central light-field image as a target  $\mathbf{Z}$ , we iterate through the 2D angular light-field where each angular view is a source  $\mathbf{R}$ .

After the color transfer, intensities undergo an automatic dynamic range alignment using the lower and upper histogram percentiles (0.005 and 99.9) of the central light-field image. This is followed by a gamma correction according to the sRGB standard. An evaluation of the herein proposed methods is carried out in Section III.

### E. LfpRefocuser

The LFPREFOCUSER enables computational change of the optical focus in light-field images by offering 3 different mechanisms. The fundamental technique is the conventional *shift-and-sum* method as originally presented by Isaksen [35] who employed a light-field taken by an array of cameras. This concept was transferred to sub-aperture images  $E_i[s_j]$  by Ng *et al.* [1], where a refocused image vector  $E'_a[s_j]$  with synthetic focus scale  $a$  is given by

$$E'_a[s_j] = \sum_{i=-c}^c E_i[s_{j-a(c+i)}], \quad a \in \mathbb{Q} \quad (37)$$

for the 1D case leaving out variables of the second spatial domain. A refinement option enables sub-pixel precision in refocusing via upsampling each sub-aperture image before integration. While this numerically increases the spatial resolution, it also allows for sub-sampling the focal depth range.

As an alternative, *shift-and-sum* is accomplished based on an aligned micro image array  $E_{fs}[s_j, u_{c+i}]$ , which writes

$$E'_a[s_j] = \sum_{i=-c}^c E_{fs}[s_{j+a(c-i)}, u_{c+i}], \quad a \in \mathbb{Q} \quad (38)$$

and can be thought of as employing an interleaved convolution kernel as shown in our previous work examining FPGA-based refocusing [36]. For an intuitive understanding of computational refocusing, the concept behind Eq. (38) is illustrated in Fig. 4 with the aid of paraxial optics. Interested readers may note that refocused distances can be pinpointed in space, using this model [2], [22].

The featured LFPSCHEIMPFLUG class mimics the identically named principle on a computational level. Tilting the refocused plane is achieved by fusing spatial image areas from a monotonically varying synthetic focus parameter  $a$ . This enables tilted focus renderings along horizontal, vertical and both diagonal image corner-to-corner directions, where focal start and end points rely on provided  $a$ . Figure 18 depicts a plenoptic-based photograph exhibiting the Scheimpflug effect.

## III. RESULTS

### A. Calibration

Our proposed generic calibration is tested in different scenarios with synthesized micro images varying in size, number, orientation angles, geometric packing and pixel noise.

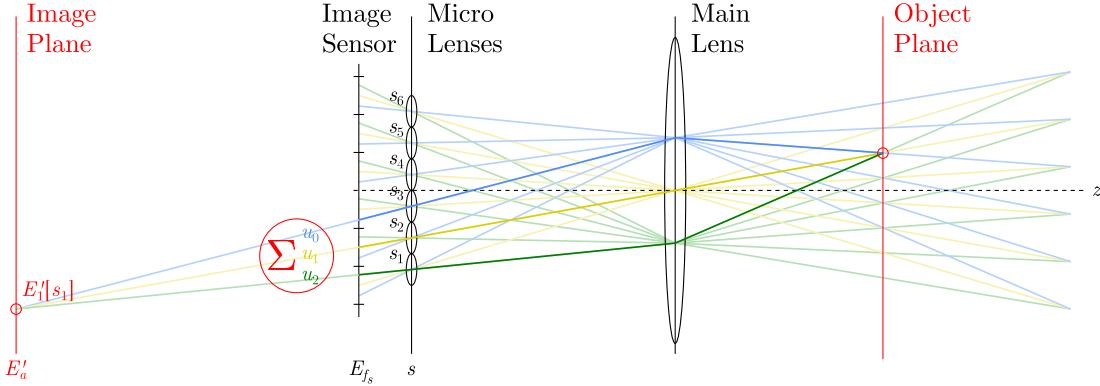


Fig. 4. Refocusing concept with chief ray intersections indicating object and image plane for a recovered point  $E'_1[s_1]$ .

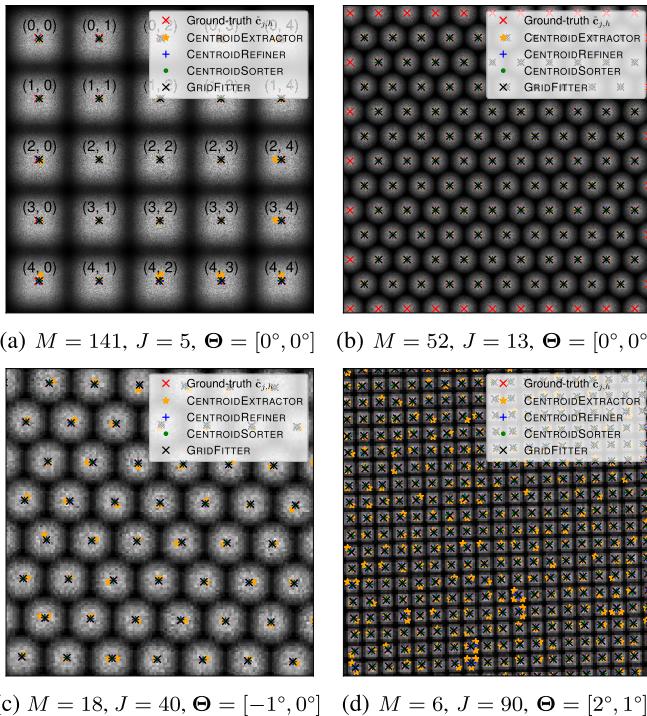


Fig. 5. Auto-calibration results from synthetic data showing noisy micro images varying in size, tilts  $\Theta = [\theta_z, \theta_x]$  about the  $z$ - $x$ -axes and packing  $\mathcal{P}$ , which is rectangular in (a); hexagonal in (b); tilted hexagonal in (c); and tilted rectangular in (d).

The validation is carried out through respective ground-truth references. Results are depicted in Fig. 5 for visual inspection.

To analyze the PITCHESTIMATOR, scale space maxima along  $\nu$  are presented from the examples shown in Fig. 5. For a quantitative examination of the centroid accuracy, we use  $C$  as a deviation metric in pixel unit given by

$$C = \frac{1}{JH} \sum_{j=1}^J \sum_{h=1}^H \|\hat{\mathbf{c}}_{j,h} - \mathbf{c}_{j,h}\|_2 \quad (39)$$

where  $\mathbf{c}_{j,h}$  represents the output of each centroid detection method and  $\hat{\mathbf{c}}_{j,h}$  acts as the ground-truth. Numerical results from Eq. (39) using examples in Fig. 5 are provided in Table I. Figure 5(a) suggests that approximates from the CENTROIDEXTRACTOR appear significantly off

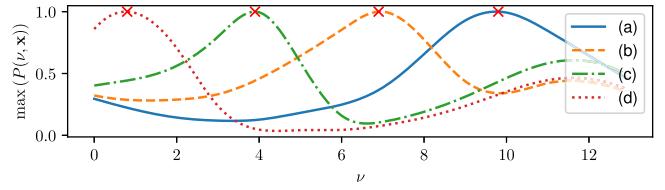


Fig. 6. Scale maxima analysis from PITCHESTIMATOR with pyramid data from Fig. 5 where crosses signify respective  $\nu^*$ .

TABLE I  
CENTROID DEVIATION  $C$  IN PIXEL UNIT

Fig. 5	Stage 1	Stage 2		Stage 3	Stage 4	Benchmark comparison				
		CEN. EXTR.	peak area			CEN. REFINER	CEN. SORTER	GRID FITTER	[12]	[36]
(a)	6.171	2.984	1.941	1.849	1.845	n.a.	n.a.	n.a.	n.a.	n.a.
(b)	0.854	0.237	0.124	0.124	0.027	0.381	n.a.	n.a.	n.a.	n.a.
(c)	1.741	0.667	0.167	0.166	0.010	0.261	1.121	0.007	n.a.	n.a.
(d)	1.162	0.419	0.311	0.297	0.007	n.a.	n.a.	n.a.	n.a.	n.a.

with regards to the ground-truth while subsequent refinement stages enhance the accuracy. Closer inspection reveals that the CENTROIDEXTRACTOR may yield several centroids for a micro image suffering from noise. This is a side effect of NMS which would have been of greater concern if white images were not convolved with  $\nabla^2 G(\sigma^*, \mathbf{x})$  prior to NMS, as this proves to cancel out false maxima. Candidates passing through LFPExtractor thus tend to be close to the ground-truth.

From Table I, it follows that the CENTROIDREFINER uses Eq. (5) as opposed to Eq. (4), with the latter propagating more imprecise coordinates to subsequent procedures. The CENTROIDSORTER yields better results, because it merges centroids belonging to the same micro image. Improvements of the GRIDFITTER arise from a large number of data points in the global centroid regression, which outperforms state-of-the-art methods [37], [38]. Due to the absence of vignetting in Fig. 5, regularization was omitted, but is shown in Fig. 7.

On closer inspection of Fig. 7, one may note that mechanical vignetting causes off-center micro images to be of non-radially symmetric shape, shifting detected centers away from their actual physical counterparts. This phenomenon has been recognized and addressed by peers [25], [37], [39], [40]. To work against this displacement in a cost-efficient manner, we have introduced a regularization term  $\beta R(\cdot)$  to the LM

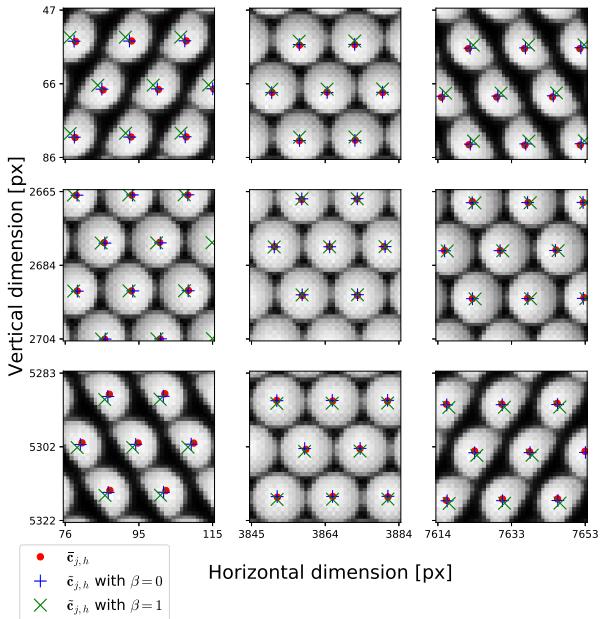


Fig. 7. **Lytro calibration** with centroids from Eq. (5) (blue), LM regression (green) and regularization where  $\beta = 1$  (red).

regression in Eq. (13) for penalization of corrupted centroid peaks. Despite the promising GRIDFITTER results, its usage is left optional to allow for calibration of inconsistent lattice spacings.

### B. Sub-Aperture Images

The central view of a light-field generally suffers the least of aberrations and thus, exposes best image quality. Figure 9 depicts decoded central views of light-field photographs from the available IRISA dataset [41] rendered by state-of-the-art plenoptic imaging pipelines for comparison. Closer inspection reveals that central views from LFTOOLBOX v0.5 appear brighter, however, fail to preserve bright image details.

For a quantitative assessment of the different pipelines, we seek a metric serving as an objective measure. Since central sub-aperture images lack ground-truth references, we employ a widely accepted no-reference-based technique named *Blind Reference-less Image Spatial Quality Evaluator* (BRISQUE) [42] from the *pybrisque* implementation [43]. Scores from the BRISQUE metric are presented in Fig. 8. Light-field denoising as borrowed by [18] was left out in the evaluation as it is considered equally effective for each pipeline and thus, not crucial for the decomposition. For our benchmark comparison, all pipelines are set to render with de-vignetting, color and sRGB options.

The BRISQUE metric is known to be sensitive to noise and blur by local statistical analysis so that an information loss caused by image processing modules would deteriorate the score. Low scores of our pipeline in Fig. 8 justify that on average our decomposition retains the physical information optimally. Our central sub-aperture images outperform the other two pipelines in 24 out of 36 cases of the dataset samples [41]. We achieve a total score of 1567 whereas CLIM-VSENSE yields 1595 and the LFTOOLBOX v0.5 gets 1702 scores.

The reason for our strong results relies on our proposed extensions to existing pipelines. This includes the treatment of detected hot-pixels prior to Bayer demosaicing to prevent outliers from propagating to adjacent pixels. We further take advantage of Menon's method [30] for debayering. In addition, the LFPRESAMPLER contributes to lower scores by conducting micro image alignment in an element-wise manner instead of re-mapping the entire plenoptic image as a whole [38]. Another reason for the improvements is that PLENOPTICAM benefits from insights previously made available by peers such as the de-saturation approach [18]. Apart from that, our automatic dynamic range alignment based on percentiles of different color spaces tends to be a more generic solution for images of various exposure, which becomes apparent by comparing (g) with (k) in Fig. 9.

For comparisons with the Lytro engine, results were computed with LYTRO DESKTOP v3.4.4 and LYTRO POWER TOOLS v1.0. Thereby, metrics indicate that the latter yields brighter images with larger field of view and allows for numerical parametrization (e.g.,  $\lambda$  for focus), making experimental reproduction easier. All Lytro results presented hereafter were therefore generated using LYTRO POWER TOOLS v1.0. Figure 10 provides a comparison of light-fields rendered by LYTRO POWER TOOLS v1.0 and PLENOPTICAM v1.0, respectively. On closer examination of the Lytro results in Figs. 10(a) and 10(c), occasional displacements emerge with image regions being removed or unnaturally shifted from their surroundings. Since Lytro's code has not been disclosed, the cause for this appearance can only be speculated. A study carried out by scientists formerly affiliated with Lytro suggests that a previously computed depth map is used to fuse a stack of high-resolution refocused frames to an all-in-focus image [44], [45], which in turn may be a starting point for high-resolution angular view synthesis. While this yields higher resolutions, unknown or erroneous disparity is propagated to subsequent stages, i.e. super-resolved views. Image region displacements in Figs. 10(a) and 10(c) thus might be due to false depth map disparities. Lytro's spatial resolution is numerically higher than the number of present micro lenses, further supporting the assumption of such pipeline design. These observations bring us to the conclusion that the Lytro engine omits direct access to classical sub-aperture images unlike other open-source toolchains [12], [18] including ours.

Technically, the user can generate sub-aperture images from Lytro's aligned ESLF images by completing their processing manually, including pixel rearrangements, adapting angular view positions and accounting for gamma as well as color correction. Given that those ESLF images are at a different processing stage and require additional resources, we compare our sub-aperture images with the scientific toolbox [12] and its extensions [18]. Nonetheless, we then contrast our refocusing results against those of the Lytro engine in Section II-E.

For further sub-aperture image analysis, we focus on the HEXCORRECTOR results where artifacts caused by hexagonal sampling are subject to removal. Figure 11 shows magnified portions of sub-aperture images with apparent fringe artifacts in Figs. 11(a) and 11(e) as well as rectified counterparts exposing smooth edges in Figs. 11(c) and 11(g). Although the

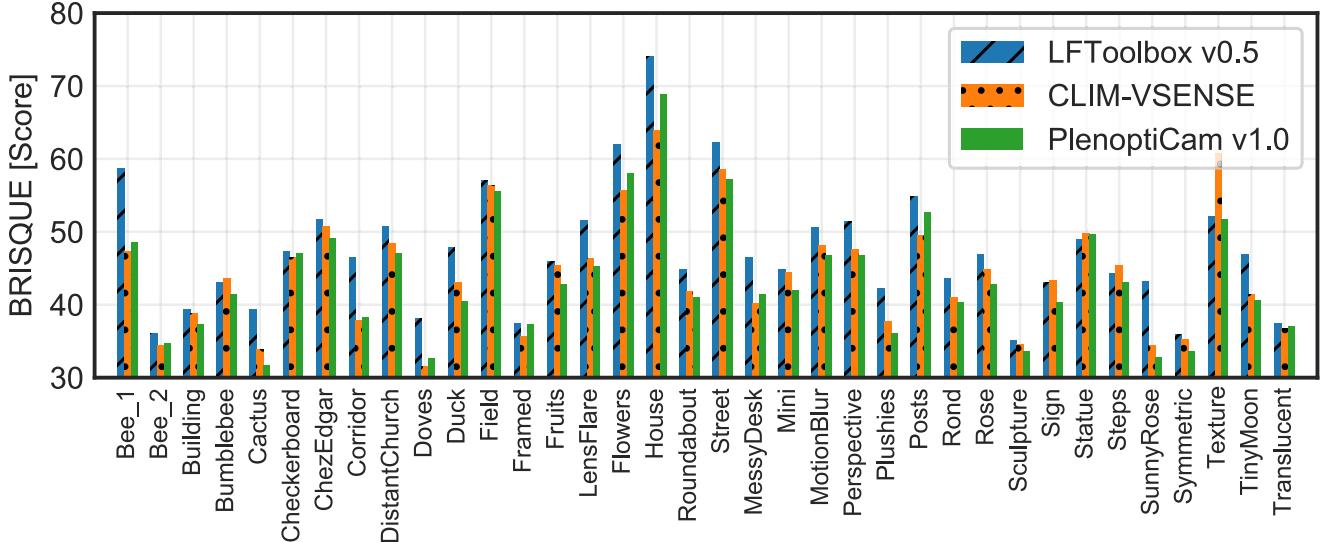


Fig. 8. Analysis of central sub-aperture images from [41] along the horizontal axis and BRISQUE metric scores along the vertical axis where lower values signify higher quality. Our tool chain outperforms others in 24 out of 36 images in total.

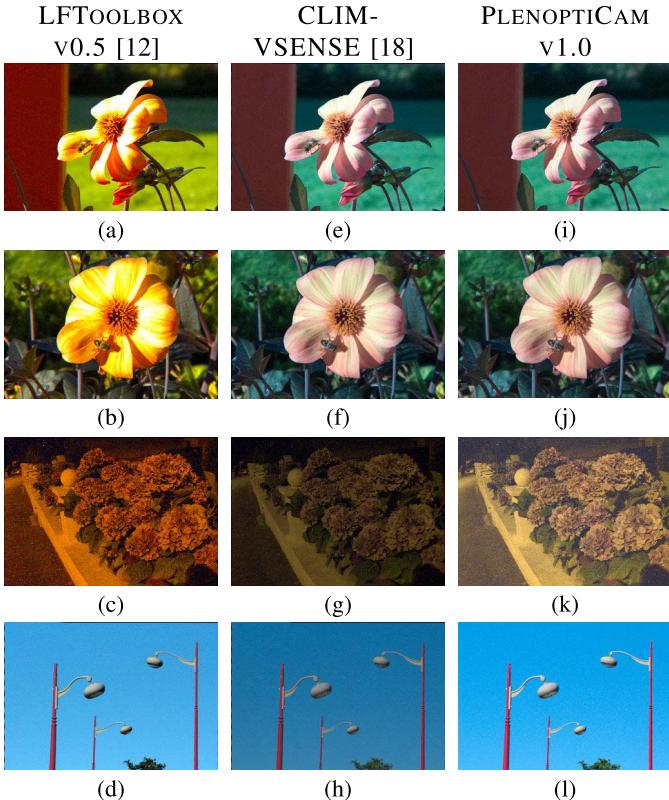


Fig. 9. Central sub-aperture images from the IRISA dataset [41] rendered in (a) to (d) by LFTOOLBOX V0.5 [12], in (e) to (h) by CLIM-VSENSE [18] and ours in (i) to (l).

efficacy of our treatment is visually notable, this improvement does not present a huge impact on BRISQUE scores. This is likely due to the blind character of the metric and its unawareness of the hexagonal sampling artifact, which can be interpreted as reasonably high spatial frequencies. Another observation made in Fig. 11 is that organic object structures

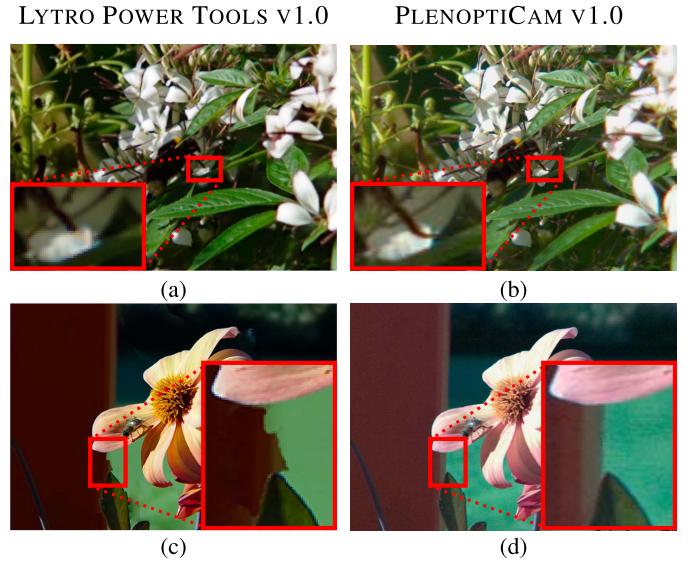


Fig. 10. View comparison with Lytro engine revealing differences between classical sub-aperture images by our pipeline in (b), (d) and Lytro's results in (a), (c), a consequence of all-in-focus rendering from depth-based segmentation [44], [45].

intentionally receive no treatment as they lack straight edges and are thus less affected.

Light-field rays arriving from non-paraxial angles generally suffer from vignetting causing light distributions to be inconsistent among views which we address with the LFP-COLOR-EQUALIZER. For visual inspection of the intensity variances, sub-aperture images are stitched together in Fig. 12 before and after treatments. A detailed analysis of sub-aperture images at a marginal position ( $i = 7, g = 2$ ) is provided in Fig. 14.

For quantitative assessment of the color transfers we compute normalized CDFs to obtain a Wasserstein metric  $W_1$  by

$$W_1 = \int_0^\infty |F(k, \mathbf{r}^{(g)}) - F(k, \mathbf{z}^{(g)})| dk \quad (40)$$

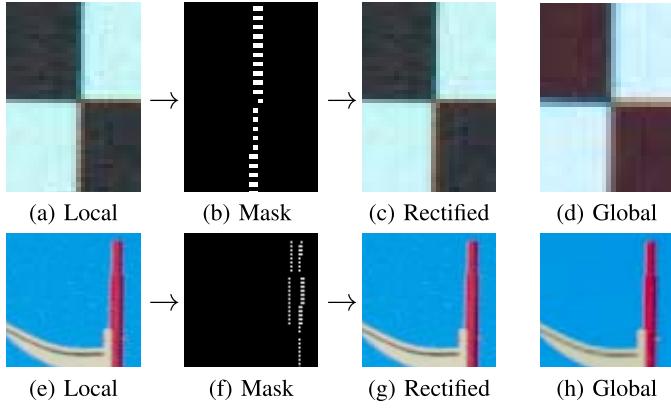


Fig. 11. **Fringe artifact reduction results** showing sub-aperture images with local resampling on the left, which exhibit fringes along vertical lines. The binary masks  $V_i[\tilde{s}_j]$  indicate pixels detected by HEXCORRECTOR that undergo rectification. The right column depicts global resampling results.

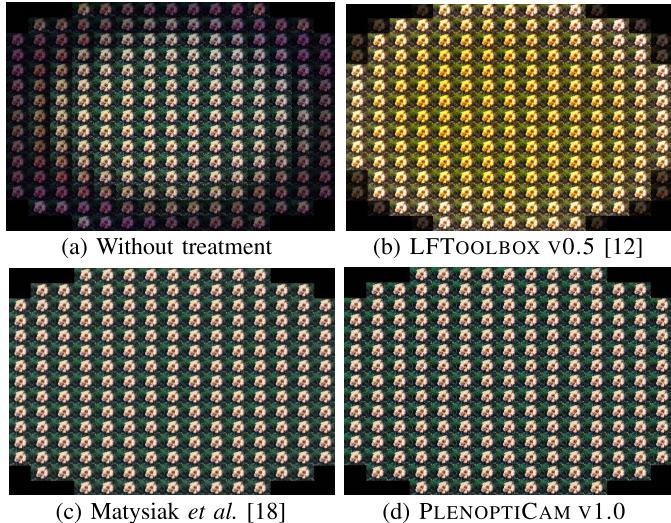


Fig. 12. **Color equalization results** with  $15 \times 15$  stitched sub-aperture images indicating illumination variances across a decomposed light-field. (a) our result without treatment, (b) Dansereau *et al.* [12], (c) Matysiak *et al.* [18] with 240 mins computation time and (d) our HM-MKL-HM compound with 81 secs computation time.

where  $\mathbf{r} \stackrel{\text{def}}{=} E_g[s_j]$  represents the marginal and  $\mathbf{z} \stackrel{\text{def}}{=} E_c[s_j]$  the central sub-aperture image serving as the reference. In addition, we employ the average histogram distance  $D_2$  by

$$D_2 = \|f(k, \mathbf{r}) - f(k, \mathbf{z})\|_2 \quad (41)$$

with  $f(k, \cdot)$  as an all-channel PDF by Eq. (28). Color consistency analysis is depicted in Fig. 13 and suggests that our low-cost transport outperforms other methods [18]–[20].

Table II provides computation times for a single Lytro Illum picture at different stages employing one physical processor of an Intel Core i7 @ 2.5 GHz. The extensive computational load imposed by Matysiak’s recoloring procedure [18] makes it impractical for us to iterate through an entire dataset.

### C. De-Vignetting

Experimental validation of our LSQ-based de-vignetting in Eqs. (19) to (22) is assessed using the Peak

TABLE II  
COMPUTATION TIME COMPARISON WHERE  $M = 15$

Process	Dansereau <i>et al.</i>	Matysiak <i>et al.</i>	Ours	
Resampling & Decoding	70 secs	125 secs	local 432 secs	global 85 secs
Hot Pixels	no support	100 secs	52 secs	
Color Eq.	no support	234 mins	81 secs	
<b>Total</b>	<b>70 secs</b>	<b>237 mins</b>	<b>565 secs</b>	<b>218 secs</b>

Signal-to-Noise-Ratio (PSNR) given by

$$\text{PSNR} = 20 \cdot \log_{10} \left( \frac{2^8 - 1}{\sqrt{\frac{1}{J} \sum_{j=1}^J (E_T[s_j] - E_c[s_j])^2}} \right) \quad (42)$$

as a metric to analyze its effectiveness. For the ground-truth data  $E_T[s_j]$ , we use a photograph divided by a white image considered noiseless. This white image is then exposed to additive Gaussian noise with  $\sigma_v = 0.15$  for de-vignetting by either classical division [12], [18] and our proposed fitting scheme. In doing so, we expect the synthetic noise to propagate during de-vignetting, however, with sufficient suppression (i.e. higher PSNR) for our LSQ approach. The results in Fig. 15 show that our method retains the image quality by gaining  $\sim 10$  dB of dynamic range with respect to noise. This is a consequence of the least-squares fit treating pixel noise as residuals. Although our evaluation contains highly amplified noise, this method is regarded equally beneficial for low-noise images.

### D. Refocusing

For the refocusing assessment, we compare the proposed LFPREFUSER module against LYTRO POWER TOOLS v1.0. Refocused images are depicted in Fig. 16. We choose the *Bumblebee.lfp* image from the IRISA dataset [41] as it exhibits organic object structures spread through a wide range of depth.

Quantitative comparison is achieved by analysis of the sharpness and BRISQUE score of local image details. For the sharpness analysis, we follow the Mavridaki and Mezaris [46] approach by transforming cropped versions of a refocused image slice  $E_a''[s_j, t_h]$  to the Fourier domain using the Discrete Fourier Transformation (DFT) and extracting the magnitude signal  $\mathcal{X}[\sigma_\omega, \rho_\psi]$  by

$$\mathcal{X}[\sigma_\omega, \rho_\psi] = \left| \sum_{j=\xi}^{\Xi} \sum_{h=\varpi}^{\Pi} E_a''[s_j, t_h] \exp(-2\pi\kappa(j\omega/(\Xi-\xi) + h\psi/(\Pi-\varpi))) \right| \quad (43)$$

where  $\kappa = \sqrt{-1}$  and  $|\cdot|$  yields absolute values. Given the 2D magnitude signal  $\mathcal{X}[\sigma_\omega, \rho_\psi]$ , the total energy  $TE$  reads

$$TE = \sum_{\omega=1}^{\Omega} \sum_{\psi=1}^{\Psi} \mathcal{X}[\sigma_\omega, \rho_\psi]^2 \quad (44)$$

where  $\Omega = \lceil (\Xi - \xi)/2 \rceil$  and  $\Psi = \lceil (\Pi - \varpi)/2 \rceil$  are borders cropping the first quarter of the unshifted magnitude signal.

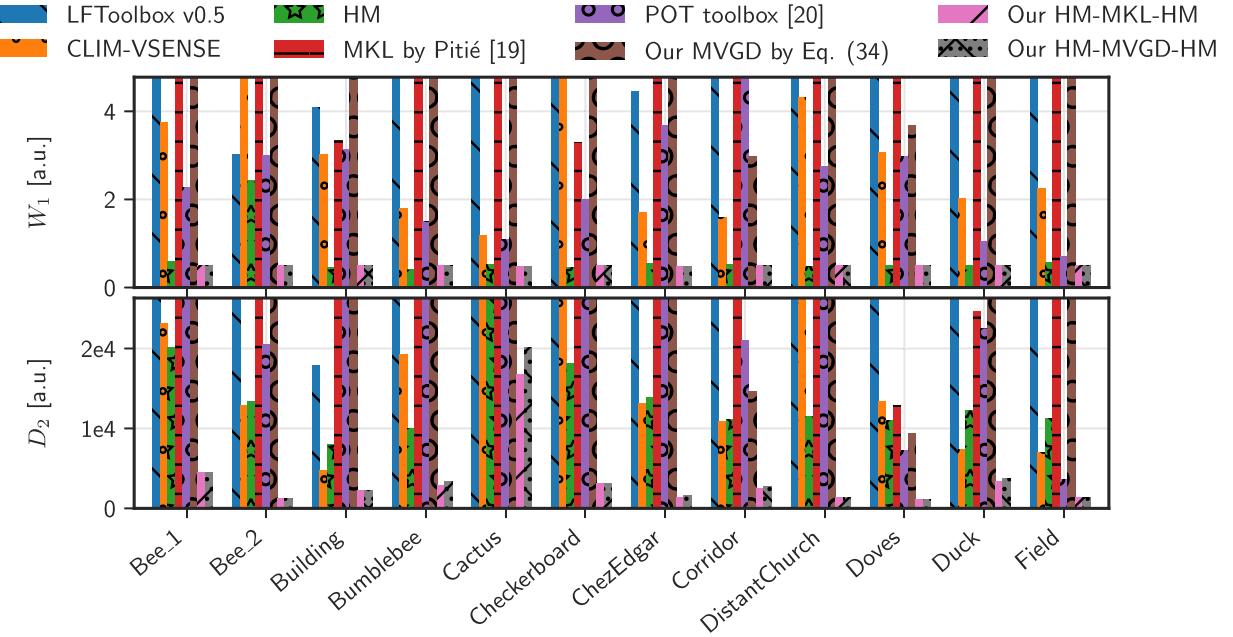


Fig. 13. **Light-field color consistency** from Wasserstein metric  $W_1$  and histogram distance  $D_2$  where low values indicate high similarity between marginal views  $E_g[s_j]$  at  $i = 7, g = 2$  and central views  $E_c[s_j]$  from a light-field dataset [41] with  $M = 15$ .

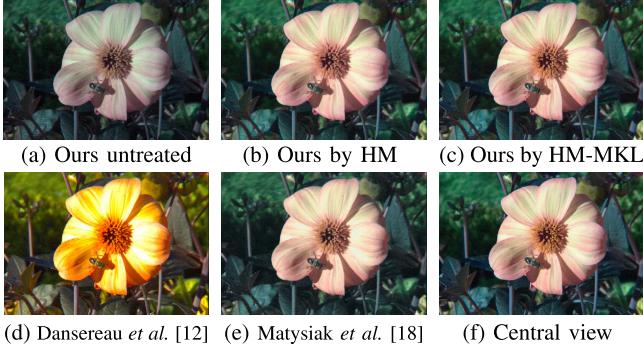


Fig. 14. **Marginal view comparison** of methods combating illumination fall-off at off-axis light-field positions using an exemplary image with  $M = 15$ . The marginal view location is  $i = 7, g = 2$  with (a) the untreated image, (b) MKL in Eq. (36) by Pitié and Kokaram [19], (c) our HM-MKL-HM compound, (d) LFTTOOLBOX V0.5 [12], (e) Matysiak et al. [18] and (f) the central sub-aperture image ( $i = 7, g = 7$ ) as a reference.

To isolate the energy of high frequency elements  $HE$ , we take the power of low frequencies and subtract it from  $TE$  reading

$$HE = TE - \sum_{\omega=1}^{Q_H} \sum_{\psi=1}^{Q_V} \mathcal{X}[\sigma_\omega, \rho_\psi]^2 \quad (45)$$

with  $Q_H$  and  $Q_V$  as scalar limits in the range of  $\{1, 2, \dots, \Omega\}$  and  $\{1, 2, \dots, \Psi\}$  separating low from high frequencies. In our experiments we set the limits to a five hundredths of the cropped image resolution. Finally, the sharpness  $S$  is a frequency ratio of refocused image portions obtained by

$$S = \frac{HE}{TE} \quad (46)$$

which serves as our blur metric.

From the cropped view of the Lytro image in Fig. 17(l), it is seen that there is no gradual decrease in blur between objects at different depth leading to an unnatural appearance

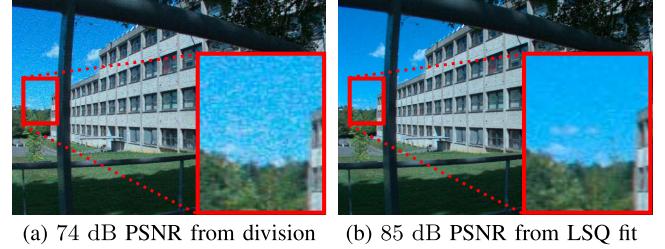


Fig. 15. **De-vignetting** from a white image with Gaussian noise  $\sigma_v = 0.15$ . The noise was absent when de-vignetting the ground-truth. Noise propagates to the light-field during division as seen in (a) and is significantly suppressed in (b) via patch-wise micro image fitting based on Eqs. (19) to (22). Our approach gains  $\sim 10$  dB of PSNR compared to the ordinary division as used by [12], [18].

of image blur. Images rendered with LFPREFOCUSER do not expose such sudden change in sharpness. It is only up to speculation on how Lytro's refocusing algorithm works. The high resolution and occasional blur artifacts around object edges suggest that Lytro super-resolves images at the first stage and then re-blurs spatial areas guided by a previously computed depth map. This may explain blur artifacts at object boundaries, which rely on the quality of a depth map.

It is therefore questionable whether the measured sharpness in Lytro's refocus algorithm arises from high frequency artifacts. By visual inspection, however, it becomes apparent that the magnified views from Lytro expose more image detail at small  $\lambda$  value. This observation is backed by respective scores from the BRISQUE metric in Table III. On the contrary, occasional artifact patterns as in Fig. 17(l) may still appeal unpleasant to human visual perception.

An exemplary result of the LFPSCHEIMPFLUG rendering engine is depicted in Fig. 18. Note how image areas in the upper left background expose sharp details, whereas the focus gradually moves toward the lower right foreground. This is

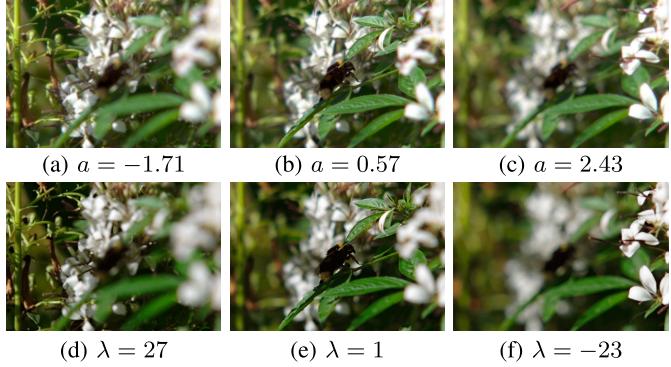


Fig. 16. **Refocused photographs** from PLENOPTICAM v1.0 in the top row with shift and sum parameter  $a$  in comparison to results processed by LYTO POWER TOOLS v1.0 in the bottom row with focus parameter  $\lambda$ .

TABLE III  
METRIC ASSESSMENT OF REFOCUSED IMAGES

Metrics	Fig. 17	(a)	(b)	(c)	(d)	(e)	(f)
Sharpness $S$		0.034	0.040	<b>0.060</b>	0.107	0.107	<b>0.224</b>
BRISQUE score		84.29	83.51	<b>81.67</b>	92.43	81.29	<b>77.29</b>

Metrics	Fig. 17	(g)	(h)	(i)	(j)	(k)	(l)
Sharpness $S$		0.107	0.107	<b>0.125</b>	0.022	0.027	<b>0.054</b>
BRISQUE score		92.43	81.29	<b>81.19</b>	94.30	90.56	<b>80.22</b>

a consequence of the synthetic focus plane being tilted with respect to the image sensor.

#### IV. CONCLUSION

Plenoptic image decomposition is a crucial task in light-field rendering for which we propose a pipeline with outstanding image quality. To the best of our knowledge, PLENOPTICAM is the only framework that allows users to automatically calibrate footage from different types of plenoptic cameras regardless of the micro lens specification. We achieve this by employing blob detection, non-maximum suppression and a generic MLA grid geometry recognition. The light-field color equalization stage is taken to a new level by using histogram matching in conjunction with an optimal transport solver to yield convincing quantitative results at reasonable computation times, which outperforms and accelerates previous methods [18]–[20]. For plenoptic image alignment, we created an alternative element-wise micro image resampling procedure. Moreover, we are first to address fringe artifacts caused by hexagonal micro lens arrangements, which are successfully reduced by our novel identification scheme. An extensive assessment of state-of-the-art sub-aperture image decomposition pipelines has been carried out in this paper with the result of outperforming other tool chains in two thirds of the cases, using a wide range of metrics such as BRISQUE, Wasserstein and histogram distance as well as a blur metric.

These findings and respective implementations are released as an open-source framework made available as a repository on GitHub [47] to allow others to participate and contribute. The herein presented framework is intended to disburden newcomers and facilitate first steps in the field of plenoptic imaging. Peers are encouraged to report bugs and actively

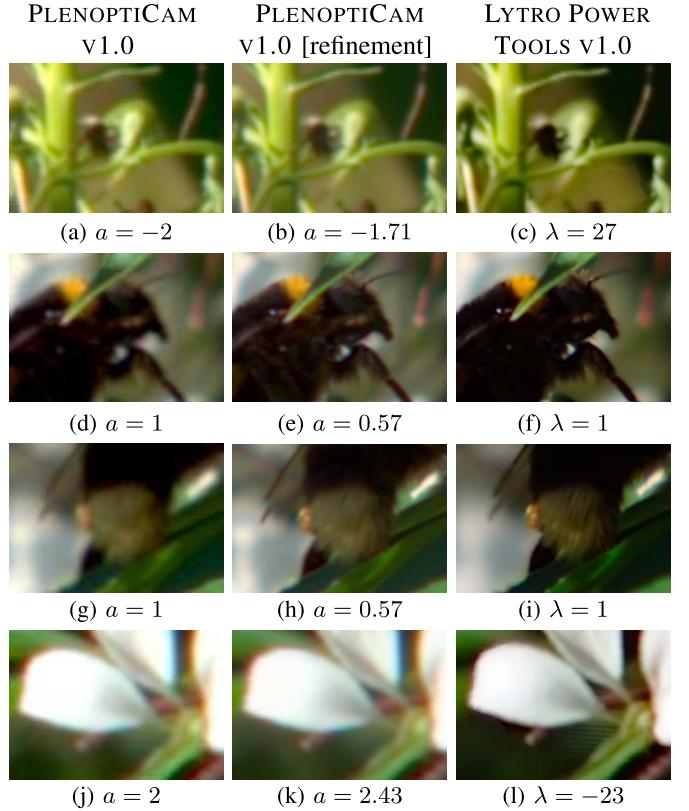


Fig. 17. **Magnified refocused image tiles** showing texture details which are quantified by our blur metric  $S$  in Table III. The left and middle column contain results from PLENOPTICAM v1.0 whereas (b), (e), (h), (k) are rendered with sub-pixel precision. The right column is obtained from LYTO POWER TOOLS v1.0 using focus parameter  $\lambda$ .

expand this software. PLENOPTICAM may lay the groundwork for future algorithm development and testing of plenoptic data.

Such extensions may include, but are not limited to, an algorithm development for the focused plenoptic camera as initially discovered by Lumsdaine and Georgiev [48]. An implementation of a super-resolution technique may be accomplished on the basis of early pioneering work [44], [49]. It should also be feasible to enhance the image quality by combating optical aberrations as shown in previous studies [45], [50]–[52] or by making use of the Bayer demosaicing similar to what Yu and Yu have revealed [53]. Until now, rectification for lens distortions exceeded the scope as this can be accomplished using traditional computer vision libraries. However, correct disparity estimation requires counteracting distortions in sub-aperture images at a preceding stage.

We believe that our work will be a substantial contribution to the field of plenoptic cameras, not limited to the image processing community. As Dansereau's LFToolbox has shown, there is a large group of researchers experimenting with plenoptic cameras demonstrating the demand for an easy-to-use software addressing the special requirements for plenoptic imaging. PLENOPTICAM may serve this broad user group including image scientists, programmers in the fields of data science, medical and industrial engineering or photography independent of the user's operating system. In particular, PLENOPTICAM facilitates light-field data preparation for visual machine learning systems such as convolutional neural



Fig. 18. **Scheimpflug rendered photograph** from LFPSCHEIMPFLUG class with  $a = \{0, 1\}$  and sub-pixel precision.

networks, which currently receive an increasing interest from a variety of scientific and industrial communities. The chosen open source license model enables cost-free usage and code republication with modifications, giving users the opportunity to extend the presented software.

## APPENDIX A SCALE SPACE THEOREM

**Theorem M** and  $\sigma^*$  are equal up to scale in Eqs. (1) and (3).

*Proof:* Let  $G(\sigma, \mathbf{x})$  be a Gaussian function of  $\mathbf{x} = [k \ l]^T$  and  $\sigma$  be the scale in a Laplacian pyramid with blob radius  $r$ , then

$$\begin{aligned} G(\sigma, k, l) &= \exp\left(-\frac{k^2 + l^2}{2\sigma^2}\right) \\ \partial_k G(\sigma, k, l) &= -\frac{k}{\sigma^2} \exp\left(-\frac{k^2 + l^2}{2\sigma^2}\right) \\ \partial_{kk} G(\sigma, k, l) &= \left(-\frac{1}{\sigma^2} + \frac{k^2}{\sigma^4}\right) \exp\left(-\frac{k^2 + l^2}{2\sigma^2}\right) \\ \nabla^2 G(\sigma, k, l) &= \partial_{kk} G(\sigma, k, l) + \partial_{ll} G(\sigma, k, l) \\ &= \left(\frac{k^2 + l^2}{\sigma^4} - \frac{2}{\sigma^2}\right) \exp\left(-\frac{k^2 + l^2}{2\sigma^2}\right) \\ r^2 &= k^2 + l^2 \\ 0 &= \frac{r^2}{\sigma^4} - \frac{2}{\sigma^2} \\ r &= \sqrt{2}\sigma \\ M &\approx 2r = 2\sqrt{2}\sigma \end{aligned}$$

□

## REFERENCES

- [1] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Stanford Univ., Stanford, CA, USA, Tech. Rep. CTSR 2005-02, 2005.
- [2] C. Hahne, A. Aggoun, V. Velisavljevic, S. Fiebig, and M. Pesch, "Refocusing distance of a standard plenoptic camera," *Opt. Exp.*, vol. 24, no. 19, pp. 21521–21540, 2016. [Online]. Available: <https://www.opticsexpress.org/abstract.cfm?URL=oe-24-19-21521>
- [3] R. Prevedel *et al.*, "Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy," *Nature Methods*, vol. 11, no. 7, pp. 727–730, 2014.
- [4] H. Li *et al.*, "Fast, volumetric live-cell imaging using high-resolution light-field microscopy," *Biomed. Opt. Exp.*, vol. 10, no. 1, p. 29, Jan. 2019. [Online]. Available: <https://www.biorxiv.org/content/early/2018/10/10/439315>
- [5] N. Bedard *et al.*, "Light field otoscope design for 3D *in vivo* imaging of the middle ear," *Biomed. Opt. Exp.*, vol. 8, no. 1, p. 260, 2017.
- [6] D. W. Palmer *et al.*, "Glare-free retinal imaging using a portable light field fundus camera," *Biomed. Opt. Exp.*, vol. 9, no. 7, p. 3178, 2018.
- [7] P. P. Srinivasan, T. Wang, A. Sreelal, R. Ramamoorthi, and R. Ng, "Learning to synthesize a 4D RGBD light field from a single image," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2243–2251.
- [8] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. 23rd Annu. Conf. Comput. Graph. Interact. Techn.*, 1996, pp. 31–42.
- [9] E. H. Adelson and J. Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 99–106, Feb. 1992.
- [10] C. Hahne, A. Aggoun, V. Velisavljevic, S. Fiebig, and M. Pesch, "Baseline and triangulation geometry in a standard plenoptic camera," *Int. J. Comput. Vis.*, vol. 126, no. 1, pp. 21–35, Jan. 2018, doi: [10.1007/s11263-017-1036-4](https://doi.org/10.1007/s11263-017-1036-4).
- [11] D. Cho, M. Lee, S. Kim, and Y.-W. Tai, "Modeling the calibration pipeline of the lytro camera for high quality light-field image reconstruction," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 3280–3287.
- [12] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 1027–1034.
- [13] N. Patel. (2011). *Lfptools*. [Online]. Available: <https://github.com/nrpatel/lfptools>
- [14] B. Esfahbod. (2013). *Python-LFP-Reader*. [Online]. Available: <https://github.com/behnam/python-lfp-reader>
- [15] Y. Bok, H.-G. Jeon, and I. S. Kweon, "Geometric calibration of micro-lens-based light-field cameras using line features," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 47–61.
- [16] M. L. Pendu and A. Smolic, "High resolution light field recovery with Fourier disparity layer completion, demosaicing, and super-resolution," in *Proc. IEEE Int. Conf. Comput. Photogr.*, Apr. 2020, pp. 1–12.
- [17] P. David, M. L. Pendu, and C. Guillemot, "White lenslet image guided demosaicing for plenoptic cameras," in *Proc. IEEE 19th Int. Workshop Multimedia Signal Process. (MMSP)*, Oct. 2017, pp. 1–6.
- [18] P. Matysiak, M. Grogan, M. L. Pendu, M. Alain, and A. Smolic, "A pipeline for lenslet light field quality enhancement," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, doi: [10.1109/ICIP.2018.8451544](https://doi.org/10.1109/ICIP.2018.8451544).
- [19] F. Pitie and A. Kokaram, "The linear monge-kantorovitch linear colour mapping for example-based colour transfer," in *Proc. IET 4th Eur. Conf. Vis. Media Prod. (CVMP)*, Nov. 2007, p. 23.
- [20] M. Perrot, N. Courty, R. Flamary, and A. Habrard, "Mapping estimation for discrete optimal transport," in *Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates, 2016, pp. 4204–4212.
- [21] C. Perwass and L. Wietzke, "Single lens 3D-camera with extended depth-of-field," *Proc. SPIE*, vol. 8291, Feb. 2012, Art. no. 829108, doi: [10.1117/12.909882](https://doi.org/10.1117/12.909882).
- [22] C. Hahne and A. Aggoun, "Plenoptisign: An optical design tool for plenoptic imaging," *SoftwareX*, vol. 10, Jul. 2019, Art. no. 100259. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352711019300159>
- [23] T. Lindeberg, *Scale-Space Theory in Computer Vision*. New York, NY, USA: Springer, 1994.
- [24] J. L. Crowley, O. Riff, and J. H. Piater, "Fast computation of characteristic scale using a half-octave pyramid," in *Proc. 4th Int. Conf. Scale-Space Theories Comput.*, Isle Skye, Scotland, 2002, pp. 1–8.
- [25] C.-K. Liang and Z. Wang, "Calibration of light-field camera geometry via robust fitting," U.S. Patent 9 420 276 B2, Aug. 16, 2016.
- [26] T. Q. Pham, "Non-maximum suppression using fewer than two comparisons per pixel," in *Advanced Concepts for Intelligent Vision Systems*, J. Blanc-Talon, D. Bone, W. Philips, D. Popescu, and P. Scheunders, Eds. Berlin, Germany: Springer, 2010, pp. 438–451.
- [27] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [28] J. J. Moré, "The Levenberg-Marquardt algorithm: Implementation and theory," in *Numerical Analysis*, G. A. Watson, Ed. Berlin, Germany: Springer, 1978, pp. 105–116.
- [29] A. Kordecki, H. Palus, and A. Bal, "Practical vignetting correction method for digital camera with measurement of surface luminance distribution," *Signal, Image Video Process.*, vol. 10, pp. 1417–1424, Nov. 2016.

- [30] D. Menon, S. Andriani, and G. Calvagno, "Demosaicing with directional filtering and a posteriori decision," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 132–141, Jan. 2007. [Online]. Available: <https://ieeexplore.ieee.org/document/4032820/>
- [31] M. Seifi, N. Sabater, V. Drazic, and P. Perez, "Disparity-guided demosaicing of light field images," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2014, pp. 5482–5486.
- [32] Y. Li and M. Sjostrom, "Depth-assisted demosaicing for light field data in layered object space," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 3746–3750.
- [33] R. Gonzalez and R. Woods, *Digital Image Processing, Global Edition*. London, U.K.: Pearson, 2017.
- [34] F. Pitie, A. Kokaram, and R. Dahyot, *Image Processing Series* (Enhancement of Digital Photographs Using Color Transfer Techniques). Boca Raton, FL, USA: CRC Press, Sep. 2008, ch. 11, pp. 295–321, doi: [10.1201/9781420054538](https://doi.org/10.1201/9781420054538).
- [35] A. Isaksen, "Dynamically reparameterized light fields," M.S. thesis, Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, Nov. 2000.
- [36] C. Hahne, A. Lumsdaine, A. Aggoun, and V. Velisavljevic, "Real-time refocusing using an FPGA-based standard plenoptic camera," *IEEE Trans. Ind. Electron.*, vol. 65, no. 12, pp. 9757–9766, Dec. 2018.
- [37] M. Schambach and F. P. Puente Leon, "Microlens array grid estimation, light field decoding, and calibration," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 591–603, Jan. 2020.
- [38] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Linear volumetric focus for light field cameras," *ACM Trans. Graph.*, vol. 34, no. 2, pp. 15:1–15:20, Mar. 2015, doi: [10.1145/2665074](https://doi.org/10.1145/2665074).
- [39] C. Pitts, T. Knight, C. Liang, and Y. Ng, "Compensating for variation in microlens position during light-field image processing," U.S. Patent 13 774 971, Aug. 9, 2013.
- [40] L. Mignard-Debise and I. Ihrke, "A vignetting model for light field cameras with an application to light field microscopy," *IEEE Trans. Comput. Imag.*, vol. 5, no. 4, pp. 585–595, Apr. 2019. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-02263377>
- [41] R. C. Daudt and C. Guillemot, *Lytro Illum Light Field Dataset*. Accessed: Jul. 2019. [Online]. Available: <https://www.irisa.fr/temics/demos/IllumDatasetLF/index.html>
- [42] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
- [43] Bukalapak. (2020). *Pybrisque*. [Online]. Available: <https://github.com/bukalapak/pybrisque>
- [44] C.-K. Liang and R. Ramamoorthi, "A light transport framework for lenslet light field cameras," *ACM Trans. Graph.*, vol. 34, no. 2, pp. 1–19, Mar. 2015, doi: [10.1145/2665075](https://doi.org/10.1145/2665075).
- [45] L.-Y. Wei, C.-K. Liang, G. Myhre, C. Pitts, and K. Akeley, "Improving light field camera sample design with irregularity and aberration," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 1–11, Jul. 2015, doi: [10.1145/2766885](https://doi.org/10.1145/2766885).
- [46] E. Mavridaki and V. Mezaris, "No-reference blur assessment in natural images using Fourier transform and spatial pyramids," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 566–570.
- [47] C. Hahne. (2020). *Plenopticam*. [Online]. Available: <https://github.com/hahne/plenopticam>
- [48] A. Lumsdaine and T. Georgiev, "Full resolution lightfield rendering," Adobe Syst., San Jose, CA, USA, Tech. Rep. 1, Jan. 2008.
- [49] T. E. Bishop, S. Zanetti, and P. Favaro, "Light field superresolution," in *Proc. IEEE Int. Conf. Comput. Photogr. (ICCP)*, Apr. 2009, pp. 1–9.
- [50] R. Ng and P. Hanrahan, "Digital correction of lens aberrations in light field photography," in *Proc. Int. Opt. Design Conf.*, Jul. 2006, pp. 441–454.
- [51] N. Cohen *et al.*, "Enhancing the performance of the light field microscope using wavefront coding," *Opt. Exp.*, vol. 22, no. 20, pp. 24817–24839, Oct. 2014. [Online]. Available: <https://www.opticsexpress.org/abstract.cfm?URI=oe-22-20-24817>
- [52] C.-K. Liang, C. Pitts, C. Craddock, and G. Myhre, "Light-field aberration correction," U.S. Patent 9628684 B2, Apr. 18, 2017.
- [53] Z. Yu, J. Yu, A. Lumsdaine, and T. Georgiev, "An analysis of color demosaicing in plenoptic cameras," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 901–908.



**Christopher Hahne** (Member, IEEE) received the B.Sc. degree in media technology from the University of Applied Sciences Hamburg in 2012. After becoming a Visiting Student at Brunel University in 2012, he started an M.Phil. and transferred to a bursary-funded Ph.D. programme at the University of Bedfordshire together with his supervisor Prof. Amar Aggoun. He is currently a Visiting Scholar affiliated with the School of Mathematics and Computer Science, University of Wolverhampton. He was working with the Research and Development Department of Rohde & Schwarz and the ARRI Group. In his professional career, he has worked at TrinamiX GmbH (BASF SE) and Pepperl+Fuchs SE, which led to two granted patents. Upon completion of his doctoral studies, he published ten high impact peer-reviewed journal articles and conference papers, most of them written as the main author. His research interests concentrate on algorithm development in the fields of computer vision, audio processing, and depth sensing. In September 2021, he will join the ARTORG Center, Artificial Intelligence in Medical Imaging (AIMI) Group, University of Bern, as an Advanced Postdoctor.

Department of Rohde & Schwarz and the ARRI Group. In his professional career, he has worked at TrinamiX GmbH (BASF SE) and Pepperl+Fuchs SE, which led to two granted patents. Upon completion of his doctoral studies, he published ten high impact peer-reviewed journal articles and conference papers, most of them written as the main author. His research interests concentrate on algorithm development in the fields of computer vision, audio processing, and depth sensing. In September 2021, he will join the ARTORG Center, Artificial Intelligence in Medical Imaging (AIMI) Group, University of Bern, as an Advanced Postdoctor.



**Amar Aggoun** received the Ingénieur d'état degree in electronics engineering from the Ecole Nationale Polytechnique d'Alger, Algiers, Algeria, in 1986, and the Ph.D. degree in electronic engineering from the University of Nottingham. His academic career started at the University of Nottingham, where he was a Research Fellow in low power DSP architectures and a Visiting Lecturer in electronic engineering and mathematics. In 1993, he was a Lecturer with De Montfort University and progressed to the position of a Principal Lecturer in 2000. In 2005, he joined Brunel University as a Reader in information and communication technologies. From 2013 to 2016, he was the Head of the School of Computer Science and Technology, University of Bedfordshire. He was the Director of the Institute for Research in Applicable Computing, which oversees all the research within the school. He is currently the Head of the School of Mathematics and Computer Science and a Professor of visual computing with the University of Wolverhampton. His research is mainly focused on 3D imaging and immersive technologies and he successfully secured and delivered research contracts worth in excess of £6.9M, funded by the Research Councils U.K., Innovate U.K., and the European Commission and Industry. Amongst the successful project, he was the Initiator, the Principal Coordinator, and a Manager of a project sponsored by the EU-FP7 ICT-4-1.5-Networked Media and 3D Internet, namely Live Immersive Video-Audio Interactive Multimedia (3D VIVANT). He holds three filed patents, published more than 200 peer-reviewed journals and conference publications, and contributed to two white papers for the European Commission on the Future Internet. He served as an Associate Editor for the IEEE/OSA JOURNAL OF DISPLAY TECHNOLOGY.

he joined Brunel University as a Reader in information and communication technologies. From 2013 to 2016, he was the Head of the School of Computer Science and Technology, University of Bedfordshire. He was the Director of the Institute for Research in Applicable Computing, which oversees all the research within the school. He is currently the Head of the School of Mathematics and Computer Science and a Professor of visual computing with the University of Wolverhampton. His research is mainly focused on 3D imaging and immersive technologies and he successfully secured and delivered research contracts worth in excess of £6.9M, funded by the Research Councils U.K., Innovate U.K., and the European Commission and Industry. Amongst the successful project, he was the Initiator, the Principal Coordinator, and a Manager of a project sponsored by the EU-FP7 ICT-4-1.5-Networked Media and 3D Internet, namely Live Immersive Video-Audio Interactive Multimedia (3D VIVANT). He holds three filed patents, published more than 200 peer-reviewed journals and conference publications, and contributed to two white papers for the European Commission on the Future Internet. He served as an Associate Editor for the IEEE/OSA JOURNAL OF DISPLAY TECHNOLOGY.