SUPER-RESOLUTION IMAGE SYNTHESIS USING THE PHYSICAL PIXEL ARRANGEMENT OF A LIGHT FIELD CAMERA

Kazuki OHASHI, Keita TAKAHASHI, Mehrdad PANAHPOUR TEHRANI, Toshiaki FUJII

Graduate School of Engineering, Nagoya University
Department of Electrical Engineering and Computer Science

ABSTRACT

We propose a method for super-resolution image synthesis that accurately handles the physical pixel arrangement of a light field (plenoptic) camera. We use a Lytro camera to obtain 4D light field data (a set of multi-viewpoint images) through a micro-lens array. The light field data are multiplexed on a single image sensor, and thus, the data is first de-multiplexed into a set of multi-viewpoint (sub-aperture) images. However, the de-multiplexing process usually involves interpolation of the original data such as demosaicing for a color filter array and pixel resampling for the non-square micro-lens arrangement. During this interpolation, some information is added or lost to/from the original data. In contrast, our method can preserve the originally captured data as they are, and directly use them for the super-resolution image synthesis, where the super-resolved image and the corresponding depth map are alternatively refined. We experimentally demonstrate that our method can achieve higher image quality than that with a standard Light Field Toolbox.

Index Terms— light field camera, plenoptic camera, super-resolution image synthesis

1. INTRODUCTION

Light field (plenoptic) cameras [1, 2, 3, 4, 5, 6, 7, 8] have been attracting much attention in recent years thanks to their capability of easily obtaining dense 3D information with a single camera. A typical implementation for such cameras is to place a micro-lens array in front of an image sensor [3, 6, 7, 8]. An image obtained by such a camera contains 4D light field data, which is equivalent to dense multi-viewpoint images, in a mutiplexed way. Applications of dense 4D ligh field data includes digital refocusing, depth estimation, free-viewpoint image synthesis and panorama image generation [3, 9, 10, 11, 12, 13, 14, 15].

One of the important issues in light field cameras is the resolution; an image from each viewpoint has a quite limited resolution because many images from different viewpoints are multiplexed on a single image sensor. In a Lytro camera [7], the resolution of each de-multiplexed (sub-aperture) image is about 380 $\times$ 380 pixels when we use Light Field Toolbox [16]. To increase the resolution of a target image, information from other viewpoints can be exploited under the framework of super-resolution image synthesis [9, 17, 18, 10, 13]. However, no previous work accurately handled the physical pixel arrangement of a light field camera, where the micro-lens array is arranged in a hexagonal grid and a color filter array is placed on the image sensor.

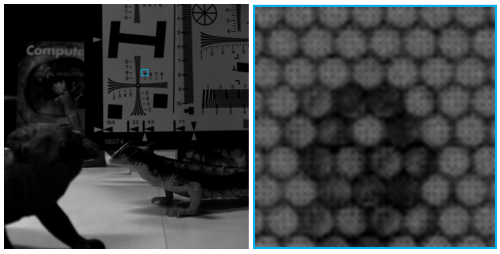

Fig. 1. Raw image and close-up

In this paper, we propose a method of super-resolution image synthesis for a Lytro camera, which handles its physical pixel arrangement as accurately as possible. When creating a set of de-multiplexed (sub-aperture) images, we preserve the originally captured information as it is, similar to [14], instead of using a standard interpolation-based method such as Light Field Toolbox [16]. Our method can work with such non-interpolated sub-aperture images, and achieve higher image quality than that with the standard Toolbox.

2. CREATION OF SUB-APERTURE IMAGES

A raw image taken by a Lytro camera and a close-up are shown in Fig. 1. As observed in the close-up, micro-lenses are arranged in a hexagonal grid. The number of micro-lenses is denoted by $l = l_x \times l_y$. A set of pixels behind each micro-lens is called a subimage.

Generally, a raw image is converted (de-multiplexed) into a set of multi-viewpoint images called sub-aperture images. A sub-aperture image is created by gathering a pixel from each subimage, which is located at the same position with respect to the center of the subimage. Consequently, the number of sub-aperture images is at most the number of pixels in each subimage. The number of pixels of a sub-aperture image ($n_x \times n_y$) is equal to the number of micro-lenses, i.e. $n_x = l_x$ and $n_y = l_y$. This creation procedure includes two problems. First, each pixel has only one color value due to a

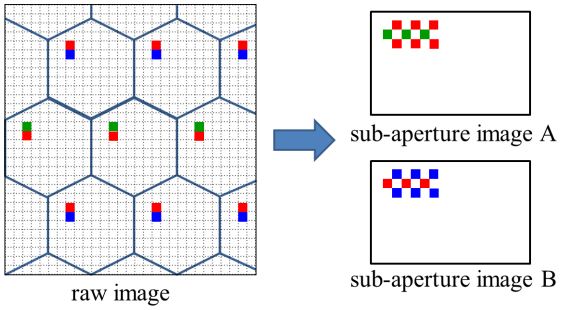

Fig. 2. Creating sub-aperture images by direct method

color filter array placed on the image sensor. Second, pixels in a sub-aperture image are arranged not in a square grid but in a hexagonal grid due to the arrangement of micro-lenses.

A typical method for creating sub-aperture images is as follows [3, 11, 12, 19, 16, 20, 15, 21]. For the first problem, a standard demosaicing algorithm is adapted to give all three color values in RGB to all of the pixels. For the second problem, pixels are resampled to make a square arrangement from the original pixels. We call this method interpolation-based method because the original information are interpolated in demosaicing and resampling steps.

Although the interpolation-based method is widely employed, some information are added and lost during the interpolation processes. Therefore, we adopt a different method [14], which is called "direct method", hereafter in this paper. As shown in Fig. 2, this method preserves the original structure of pixel arrangement in a raw image. No demosaicing is performed, and thus, each pixel has at most one color value. To keep the hexagonal arrangement of a sub-aperture image, void pixels are inserted alternatively. As a result, the width of sub-aperture image is doubled, i.e. $n_x = 2l_x$ and $n_y = l_y$. The advantage of the direct method lies on the fact that no information are added or lost to/from the original raw image. This advantage becomes significant for not only depth estimation as shown in [14], but also super-resolution image synthesis as firstly demonstrated by our study.

3. SUPER-RESOLUTION IMAGE SYNTHESIS

We propose a super-resolution image synthesis method that can accurately handle the physical pixel arrangement of a light field camera. As inputs, we have sub-aperture images written as $\mathbf{y}^{(k)} \in \mathcal{R}^n$ with $n = n_x \times n_y \times 3$ and $k \in \mathcal{K} = [1, \cdots, K]$ where K is the number of the sub-aperture images. Some elements in $\mathbf{y}^{(k)}$ are missing if we use the direct method to create it. Our goal is to estimate a higher resolution image $\mathbf{x} \in \mathcal{R}^N$ and the corresponding depth map $\mathbf{d} \in \mathcal{R}^N$, where $N = N_x \times N_y \times 3$, at the viewpoint of $\mathbf{y}^{(k_c)}$ where $k_c \in \mathcal{K}$. We employ a super-resolution image synthesis method, where $\mathbf{x}$ and $\mathbf{d}$ are jointly estimated because they depend on each other. First, $\mathbf{x}$ is initialized as $\mathbf{x}^{(0)}$. Then, depth estimation (Eq. (1)) and super-resolution image synthesis (Eq. (2)) are performed alternatively to update $\mathbf{d}$ and $\mathbf{x}$ until convergence, as will be detailed in Sections 3.1 and 3.2, respectively. For the m-th iteration, the update process is

$$\mathbf{d}^{(m)} = \arg\min_{\mathbf{d}} E_{depth}(\mathbf{x}^{(m)}, \mathbf{d}) \tag{1}$$

$$\mathbf{x}^{(m+1)} = \arg\min_{\mathbf{x}\in\mathcal{X}} E_{SR}(\mathbf{x}, \mathbf{d}^{(m)}), \tag{2}$$

where set $\mathcal{X}$ will be explained in Section 3.2. To obatin $\mathbf{x}^{(0)}$, a sub-aperture image corresponding to $\mathbf{y}^{(k_c)}$ is generated with demosaicing, and upsampled to the size of $\mathbf{x}$ by using bicubic interpolation. Note that although this joint estimation method is similar to [22, 9, 20], we are the first to use it with a careful consideration of the physical pixel arrangement of a light field camera.

3.1. Updating depth estimate

The energy function $E_{depth}(\mathbf{x}, \mathbf{d})$ is formulated as

$$E_{depth}(\mathbf{x}, \mathbf{d}) = \sum_{i\in[1,N]} \sum_{k\in\mathcal{K}\setminus k_c} C^{(k,i)}(\mathbf{x}, d_i) \tag{3}$$

where $C^{(k,i)}(\mathbf{x}, d)$ is a matching cost for the i-th pixel on $\mathbf{x}$ against the k-th image and given by

$$C^{(k,i)}(\mathbf{x}, d) = \left\| \mathbf{W}_i \, \delta^{(k)}(\mathbf{x}, d) \right\|^2. \tag{4}$$

Here, $\delta^{(k)}(\mathbf{x}, d) \in \mathcal{R}^N$ is the difference vector between $\mathbf{x}$ and $\mathbf{y}^{(k)}$ with an assumed depth d, and $\mathbf{W}_i$ is a diagonal matrix that extracts only the pixels that are spatially neighboring to the i-th pixel, which makes the matching cost being accumulated over a local window. Vector $\delta^{(k)}(\mathbf{x}, d)$ is given by

$$\delta^{(k)}(\mathbf{x}, d) = \mathbf{U}\mathbf{y}^{(k)} - \mathbf{P}^{(k)}_{N\times N}\mathbf{M}^{(k)}(d)\,\mathbf{B}\,\mathbf{x} \tag{5}$$

where $\mathbf{U} \in \mathcal{R}^{N\times n}$ is a nearest-neighbor upsampling operator, and $\mathbf{P}^{(k)}_{N\times N} \in \mathcal{R}^{N\times N}$ is a pixel arrangement matrix which is adapted to $\mathbf{U}\mathbf{y}^{(k)}$ as will be discussed later. $\mathbf{M}^{(k)}(d) \in \mathcal{R}^{N\times N}$ represents a translation where all the pixels are moved uniformly according to the depth value d. $\mathbf{B} \in \mathcal{R}^{N\times N}$ represents a point spread function to compensate the difference of pixel sizes between $\mathbf{x}$ and $\mathbf{y}^{(k)}$. It should be noted that all the matrix of Eq. (5) are implemented as image processing operators, and thus, it is unnecessary to use huge memory space for directly keeping those matrices. After Eq. (3) is minimized, a 3×3 median filter is applied to $\mathbf{d}$ to reduce isolated noises.

3.2. Updating super-resolved image

The energy function $E_{SR}(\mathbf{x}, \mathbf{d})$ is formulated as

$$E_{SR}(\mathbf{x}, \mathbf{d}) = \frac{1}{2} \sum_{k\in\mathcal{K}} \left\| \mathbf{y}^{(k)} - \mathbf{A}^{(k)}(\mathbf{d})\,\mathbf{x} \right\|^2 + \lambda\, R(\mathbf{x}) \tag{6}$$

where the first term represents an observation model, and the second term $R(\mathbf{x})$ is a regularizer to enforce the smoothness

on $\mathbf{x}$ with a nonnegative weight λ. In the observation model, $\mathbf{A}_k(\mathbf{d}) \in \mathcal{R}^{n\times N}$ is decomposed into

$$\mathbf{A}^{(k)}(\mathbf{d}) = \mathbf{P}^{(k)}_{n\times N}\mathbf{D}\,\mathbf{M}^{(k)}(\mathbf{d})\,\mathbf{B}. \tag{7}$$

Here, $\mathbf{P}^{(k)}_{n\times N} \in \mathcal{R}^{n\times N}$is a pixel arrangement matrix which is adapted to $\mathbf{y}^{(k)}$ as will be discussed later. $\mathbf{D} \in \mathcal{R}^{n\times N}$ denotes a subsampling operator. $\mathbf{M}^{(k)}(\mathbf{d}) \in \mathcal{R}^{N\times N}$ represents a pixel-wise translation where we move each pixel individually according to the depth map $\mathbf{d}$ considering occlusions among the pixels. $\mathbf{B} \in \mathcal{R}^{N\times N}$ represents a point spread function similar to $\mathbf{B}$ in Eq. (5).

When minimizing Eq. (6) using current image $\mathbf{x}^{(m)}$ and depth map $\mathbf{d}^{(m)}$, we limit the solution space of $\mathbf{x}^{(m+1)}$ to

$$\mathcal{X} = \left\{\mathbf{x} \mid \mathbf{x} = \mathbf{x}^{(m)} - \alpha\nabla E_{SR}(\mathbf{x}^{(m)}, \mathbf{d}^{(m)}),\, \alpha \in \mathcal{R}\right\}. \tag{8}$$

It is equivalent to performing a single step update of gradient decent method. Hereafter, we omit $\mathbf{d}^{(m)}$, which is constant during this procedure, to simplify expressions. Specifically, the current high resolution image $\mathbf{x}^{(m)}$ is updated to $\mathbf{x}^{(m+1)}$ with a step size $\alpha^{(m)}$ as follows [23].

$$\mathbf{x}^{(m+1)} = \mathbf{x}^{(m)} - \alpha^{(m)}\nabla E_{SR}(\mathbf{x}^{(m)}) \tag{9}$$

$$\alpha^{(m)} = \frac{\left\|\nabla E_{SR}(\mathbf{x}^{(m)})\right\|^2}{\nabla E_{SR}(\mathbf{x}^{(m)})^{\mathrm{T}}(\nabla^2 E_{SR}(\mathbf{x}^{(m)}))\nabla E_{SR}(\mathbf{x}^{(m)})} \tag{10}$$

where the expressions ∇ and ∇^2 denote gradient and Hessian operators with respect to $\mathbf{x}$. We set $R(\mathbf{x})$ as

$$R(\mathbf{x}) = \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{L}\mathbf{x} \tag{11}$$

where $\mathbf{L} \in \mathcal{R}^{N\times N}$ represents the 4-connected Laplacian operator. This regularizer is rather simple for describing natural image properties, but easy to minimize. $\nabla E_{SR}(\mathbf{x})$ and $\nabla^2 E_{SR}$ are given by

$$\nabla E_{SR}(\mathbf{x}) = -\sum_{k\in\mathcal{K}}\{\mathbf{A}^{(k)}(\mathbf{d})\}^{\mathrm{T}}(\mathbf{y}^{(k)} - \mathbf{A}^{(k)}(\mathbf{d})\mathbf{x}) + \lambda\,\mathbf{L}\mathbf{x} \tag{12}$$

$$\nabla^2 E_{SR} = \sum_{k\in\mathcal{K}}\{\mathbf{A}^{(k)}(\mathbf{d})\}^{\mathrm{T}}\mathbf{A}^{(k)}(\mathbf{d}) + \lambda\,\mathbf{L} \tag{13}$$

In the calculation of Eqs. (12) and (13), $\mathbf{A}^{(k)}(\mathbf{d})$ and $\mathbf{L}$ are implemented as image processing operators, and thus, they do not need to be represented as $N \times N$ matrices.

3.3. The role of pixel arrangement matrix P

Typically, an image is represented as a data format where each pixel has all RGB values. However, each pixel on a sub-aperture image generated by the direct method has only one color value or nothing as was discussed in section 2. To bridge the two different data formats, we use pixel arrangement matrices $\mathbf{P}^{(k)}_{N\times N}$ and $\mathbf{P}^{(k)}_{n\times N}$ in Eqs. (5) and (7). These matrices are diagonal and each diagonal element takes 1 if the corresponding value is present in the sub-aperture image or 0 otherwise. In this way, we can appropriately handle non-existing pixel values caused by the pixel arrangement structure of a light field camera. If sub-aperture images are generated by the interpolated-based method, all the diagonal elements are 1, and thus, $\mathbf{P}^{(k)}_{N\times N}$ and $\mathbf{P}^{(k)}_{n\times N}$ become identity matrices.

4. EXPERIMENTS

We captured three raw images, A, B, and C, by a Lytro camera, with the resolution of 3280 $\times$ 3280 pixels and 328 $\times$ 379 micro-lens. The size of sub-aperture images was 656 $\times$ 379 pixels with the direct method and 379 $\times$ 379 pixels with the interpolation-based method. As the interpolation-base method, we used Light Field Toolbox v0.3 [16], which is optionally combined with rectification. We used 5 $\times$ 5 sub-aperture images to super-resolve the central image to the size of 1080 $\times$ 1080 pixels. We implemented the algorithm described in section 3, which can accept both types of sub-aperture images. λ in Eq. (6) was set to 0.3. A 3 $\times$ 3 median filter is optionally applied to the final super-resolved images if it improves the quality.

Figure 3 shows the initial images and depth maps ($\mathbf{x}^{(0)}$ and $\mathbf{d}^{(0)}$) and the final images and depth maps produced by the proposed method. The final images are considerably improved compared to the initial images thanks to our proposed super-resolution image synthesis using the physical pixel arrangement (it is recommended to zoom up the images on the electric version to see details). The depth maps are also updated along the images. Although some areas without textures tend to have inaccurate depth values, they did not seriously affect the super-resolution image synthesis.

Finally, in Fig. 4, we present several close-ups of the high-resolution images obtained from the raw images, A, B, and C. We compare four scenarios as follows: (i) toolbox w/o rectification + bicubic upsampling, (ii) toolbox w/o rectification + super-resolution, (iii) toolbox w/ rectification + super-resolution, and (iv) direct + super-resolution (proposed). Obviously, the proposed method achieved the best results. This is because the proposed method accurately handles the physical pixel arrangement of a Lytro camera for super-resolution image synthesis.

5. CONCLUSIONS

We developed a framework of super-resolution image synthesis for a light field camera, where the physical pixel arrangement (the color filter array and hexagonal micro-lens arrangement) can be handled accurately. We experimentally demonstrated that our method can achieve higher image quality than that using the standard Light Field Toolbox [16]. Future work includes exploration of more suitable regularizers for the super-resolution image synthesis, which can balance computational cost and image quality.

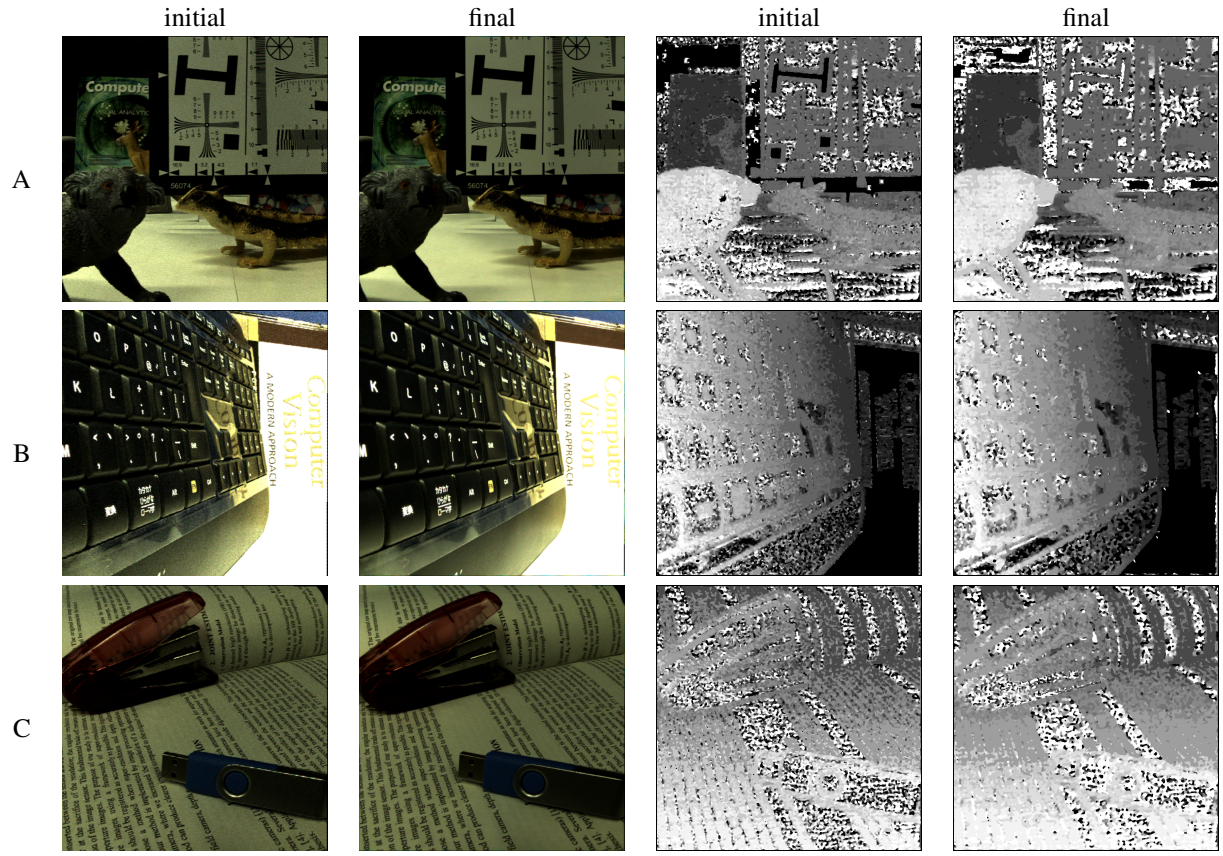

Fig. 3. The proposed method started with initial image (1st column) and depth map (3rd column), and finally produced a super-resolved image (2nd column) and a refined depth map (4th column).

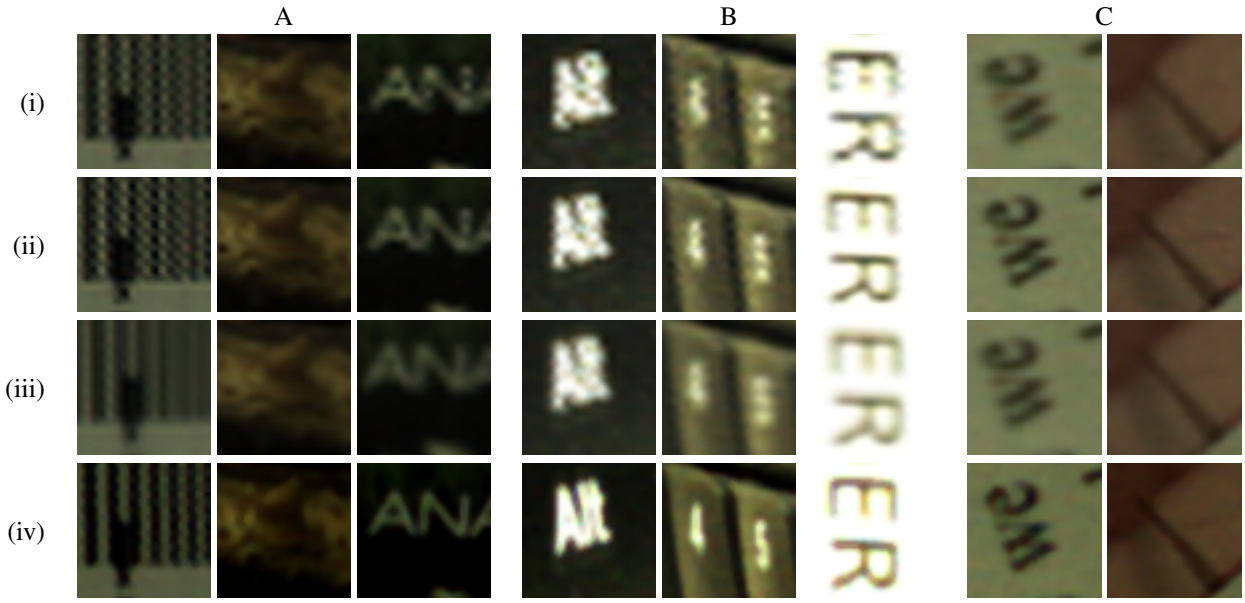

Fig. 4. Comparison of super-resolved images among four scenarios: (i) toolbox w/o rectification + bicubic upsampling, (ii) toolbox w/o rectification + super-resolution, (iii) toolbox w/ rectification + super-resolution, and (iv) direct + super-resolution (proposed).

6. REFERENCES

[1] Edward H Adelson and John Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE transactions on pattern analysis and machine intelligence*, vol. 14, no. 2, pp. 99–106, 1992.

[2] Jun Arai, Fumio Okano, Haruo Hoshino, and Ichiro Yuyama, "Gradient-index lens-array method based on real-time integral photography for three-dimensional images," *Applied optics*, vol. 37, no. 11, pp. 2034–2045, 1998.

[3] 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, 2005.

[4] Ashok Veeraraghavan, Ramesh Raskar, Amit Agrawal, Ankit Mohan, and Jack Tumblin, "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Transactions on Graphics*, vol. 26, no. 3, pp. 69, 2007.

[5] Chia-Kai Liang, Tai-Hsu Lin, Bing-Yi Wong, Chi Liu, and Homer H Chen, "Programmable aperture photography: multiplexed light field acquisition," in *ACM Transactions on Graphics (TOG)*, 2008, vol. 27, 3, pp. 55:1–55:10.

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

[7] "Lytro," `http://lytro.com/`.

[8] "Raytrix," `http://www.raytrix.de/`.

[9] F Perez Nava and JP Luke, "Simultaneous estimation of super-resolved depth and all-in-focus images from a plenoptic camera," in *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video*, 2009, pp. 1–4.

[10] Tom E Bishop and Paolo Favaro, "The light field camera: Extended depth of field, aliasing, and superresolution," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 5, pp. 972–986, 2012.

[11] Stepan Tulyakov, Tae Hee Lee, and Heechul Han, "Quadratic formulation of disparity estimation problem for light-field camera," in *ICIP*, 2013, pp. 2063–2067.

[12] Michael W Tao, Sunil Hadap, Jitendra Malik, and Ravi Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *Computer Vision (ICCV), IEEE International Conference on*, 2013, pp. 673–680.

[13] 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, 2013.

[14] Neus Sabater, Mozhdeh Seifi, Valter Drazic, Gustavo Sandri, and Patrick Perez, "Accurate disparity estimation for plenoptic images," in *Computer Vision–ECCV Workshop*, 2014.

[15] Juliet Fiss, Brian Curless, and Richard Szeliski, "Refocusing plenoptic images using depth-adaptive splatting," in *Computational Photography (ICCP), IEEE International Conference on*, 2014, pp. 1–9.

[16] 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), IEEE Conference on*, 2013, pp. 1027–1034.

[17] Todor Georgiev, Georgi Chunev, and Andrew Lumsdaine, "Superresolution with the focused plenoptic camera," in *IS&T/SPIE Electronic Imaging*, 2011, pp. 78730X–78730X.

[18] F Perez, A Perez, M Rodriguez, and E Magdaleno, "Fourier slice super-resolution in plenoptic cameras," in *Computational Photography (ICCP), IEEE International Conference on*, 2012, pp. 1–11.

[19] Donghyeon Cho, Minhaeng Lee, Sunyeong Kim, and Yu-Wing Tai, "Modeling the calibration pipeline of the lytro camera for high quality light-field image reconstruction," in *Computer Vision (ICCV), IEEE International Conference on*, 2013, pp. 3280–3287.

[20] Kazuki Ohashi, Keita Takahashi, and Toshiaki Fujii, "Joint estimation of high resolution images and depth maps from light field cameras," in *IS&T/SPIE Electronic Imaging*, 2014, pp. 90111B–90111B.

[21] Yunsu Bok, Hae-Gon Jeon, and In So Kweon, "Geometric calibration of micro-lens-based light-field cameras using line features," in *Computer Vision–ECCV*, 2014, pp. 47–61.

[22] Russell C Hardie, Kenneth J Barnard, and Ernest E Armstrong, "Joint map registration and high-resolution image estimation using a sequence of undersampled images," *Image Processing, IEEE Transactions on*, vol. 6, no. 12, pp. 1621–1633, 1997.

[23] Keita Takahashi and Takeshi Naemura, "Super-resolved free-viewpoint image synthesis based on view-dependent depth estimation," *IPSJ Transactions on Computer Vision and Applications*, vol. 7, no. 4, pp. 1529–1543, 2012.