Optics Letters

Super-resolution in a defocused plenoptic camera: a wave-optics-based approach

ERDEM SAHIN,* VLADIMIR KATKOVNIK, AND ATANAS GOTCHEV

Department of Signal Processing, Tampere University of Technology, 33720 Tampere, Finland
**Corresponding author: erdem.sahin@tut.fi*

Received 1 December 2015; accepted 26 January 2016; posted 1 February 2016 (Doc. ID 254762); published 26 February 2016

Plenoptic cameras enable the capture of a light field with a single device. However, with traditional light field rendering procedures, they can provide only low-resolution two-dimensional images. Super-resolution is considered to overcome this drawback. In this study, we present a super-resolution method for the defocused plenoptic camera (Plenoptic 1.0), where the imaging system is modeled using wave optics principles and utilizing low-resolution depth information of the scene. We are particularly interested in super-resolution of in-focus and near in-focus scene regions, which constitute the most challenging cases. The simulation results show that the employed wave-optics model makes super-resolution possible for such regions as long as sufficiently accurate depth information is available.

OCIS codes: (100.6640) Superresolution; (110.6880) Three-dimensional image acquisition; (100.6890) Three-dimensional image processing; (070.2580) Paraxial wave optics.

http://dx.doi.org/10.1364/OL.41.000998

Since the beginning of 20th century when Lippmann [1] proposed the use of an array of small lenses to capture a three-dimensional (3D) photograph, significant improvements have been achieved in 3D photography. A few decades after the first digital plenoptic camera was proposed by Adelson and Wang [2], it has been practically demonstrated that it is possible to capture a 3D photograph (light field) by a hand-held camera which then enables, for example, digital refocusing and estimating the scene depth by post-processing of the sensed light field [3,4].

The common goal of plenoptic cameras is to get angular radiance information of the scene which is lost in conventional two-dimensional (2D) photography. Usually a microlens array is inserted between the main lens and the sensor of the camera to accomplish this goal. The angular radiance information comes, however, at the expense of reduced spatial resolution. Therefore, the spatial super-resolution [5] is of particular importance in plenoptic imaging.

Super-resolution framework has been applied to both focused and defocused setups. Super-resolution of in-focus scene regions, i.e., regions which are focused onto the microlens array plane by the main lens, for the defocused setup has been identified to be the most challenging case [6,7]. As one way of coping with this difficult case, instead of ray optics model as used in [6], the more accurate wave optics model of the plenoptic imaging system has been employed in [8–10], and successful results have been reported. In those works, however, only a 2D planar object at known depth has been considered. Thus, the results have remained ambiguous for more realistic scenarios where the scene has depth variations and the knowledge of this information is limited. In the microscopic regime, Broxton *et al.* [11] have employed a similar wave optics model for the super-resolution of mostly transparent volumetric scenes, which are described by a set of voxels (volumetric elements). Transparent scene assumption is usually invalid for photographic scenes, therefore voxel-based description is inapplicable for such scenes. The super-resolution method we present here also makes use of a wave optics model of the plenoptic imaging system. The scene visible to the camera is treated as a 2D curved surface. It is then approximated with the knowledge of a low-resolution depth map. We consider realistic scenarios where the depth information can be provided with limited accuracy. We demonstrate super-resolved images for in-focus and near in-focus regions of 3D scenes and present results from which the depth accuracies required for reasonable reconstruction qualities become evident.

In the so-called defocused setup, the main lens focuses the scene onto the microlens array plane and the microlenses focus the principal plane of the main lens onto the sensor plane, i.e., according to Fig. 1, $1/f_1 = 1/z_f + 1/l_1$ and $1/f_2 = 1/l_1 + 1/l_2$, where f_1 and f_2 are the focal lengths of the main lens and microlenses, respectively. For a defocused plenoptic camera with a sensor of $N \times N$ pixels and a microlens array of $M \times M$ microlenses, the super-resolution problem can be posed as estimation of the high-resolution $(N \times N)$ image of the scene given $N/M \times N/M$ low-resolution $(M \times M)$ sub-aperture images that correspond to $N/M \times N/M$ different viewpoints on the main lens. An illustration of the considered defocused plenoptic imaging system is shown in Fig. 1 for $N = 20$ and $M = 5$.

The low-resolution depth map required for our approach can be obtained by post-processing the captured plenoptic image via a multiview depth estimation algorithm. Here, we refer

the reader to the recent work [12] for a detailed discussion on depth estimation, in particular, for plenoptic cameras. The multiview depth estimation algorithms usually provide a disparity map which gives pixel correspondences between the view of interest and a reference view. For a defocused plenoptic camera, the disparity values can be mapped to depth values by the relation $z^{i\Delta} = \{1/f_1 - [(i\Delta T_2 + B)/(Bl_1)]\}^{-1}$, where B is the baseline between those two views (i.e., the distance between adjacent sub-aperture views), T_1 and T_2 are the aperture sizes of the main lens and microlenses, respectively, and Δ and $i\Delta$ are the disparity step and the disparity value to be mapped (both in pixels), respectively. The discretization in disparity values is necessary to obtain reliable estimates. Depth values can, thus, be reliably provided only with a limited accuracy. Figure 1 illustrates the approximation of the 2D scene surface using the high-resolution depth map (circles) which is obtained by linearly interpolating such a discretized and low-resolution depth map (solid circles) given for the center view.

The high-resolution texture samples on the approximating surface correspond to the desired high-resolution image, and they are related to the captured image data through the point spread function (PSF) of the plenoptic imaging system. We make use of the PSF which is derived employing the scalar wave optics, in particular the Fresnel diffraction model. There are mainly three steps involved in the derivation of the PSF. First, taking into account the coordinate and parameter definitions in Fig. 1, for a scene illuminated by a monochromatic coherent light of wavelength λ, the field distribution just before the (s, t) plane due to a point source at $\boldsymbol{x_p} = [x_p, y_p, z_p]^T$ is found as

$$g_1(s,t;\boldsymbol{x_p}) = \frac{1}{\lambda^2 z_p l_1}\exp\left[jk\left(z_p + l_1 + \frac{x_p^2 + y_p^2}{2z_p}\right)\right] \times \exp\left[\frac{jk}{2l_1}(s^2+t^2)\right] g_1'\left(\frac{s-x_p'}{\lambda l_1}, \frac{t-y_p'}{\lambda l_1}\right), \quad (1)$$

where

$$g_1'(f_x, f_y) = \mathcal{F}\left\{P_1(x,y)\exp\left[\frac{jk}{2}\left(\frac{1}{z_p}+\frac{1}{l_1}-\frac{1}{f_1}\right)(x^2+y^2)\right]\right\}, \quad (2)$$

$\mathcal{F}\{\}$ is the Fourier transform operator, P_1 is the pupil function of the main lens, $k = 2\pi/\lambda$, and $x_p' = -l_1 x_p/z_p, y_p' = -l_1 y_p/z_p$ are the reduced coordinates [13]. In the second step, the corresponding complex field distribution on the image plane, i.e., the coherent PSF is obtained by superposing the contributions of each of the microlenses as

$$\mathrm{PSF}_c(u,v;\boldsymbol{x_p}) = \frac{1}{j\lambda l_2}\exp(jkl_2)\exp\left[\frac{jk}{2l_2}(u^2+v^2)\right] \times \sum_m\sum_n \exp\left\{\frac{-jk}{2f_2}[(mT_2)^2+(nT_2)^2]\right\} \times g_2^{mn}\left(\frac{u}{\lambda l_2}, \frac{v}{\lambda l_2}; \boldsymbol{x_p}\right), \quad (3)$$

where

$$g_2^{mn}(f_s, f_t; \boldsymbol{x_p}) = \mathcal{F}\left\{P_2(s-mT_2, t-nT_2)\exp\left[\frac{jkT_2}{f_2}(sm+tn)\right] \times \exp\left[\frac{jk}{2}\left(\frac{1}{l_2}-\frac{1}{f_2}\right)(s^2+t^2)\right] g_1(s,t;\boldsymbol{x_p})\right\}, \quad (4)$$

P_2 is the pupil function of the microlenses, and m, n are the indices for the microlenses [8]. The contributions of microlenses are found by multiplying their transmittance functions by $g_1(s, t; \boldsymbol{x_p})$, and then applying the Fresnel propagation operation by distance l_2. Since we are interested in scenes illuminated with incoherent light, the final step is to obtain the incoherent PSF. Under perfectly incoherent light assumption, the incoherent PSF is obtained by taking the magnitude-squared of the coherent PSF [13]. Thus,

$$\mathrm{PSF}(u,v;\boldsymbol{x_p}) = |\mathrm{PSF}_c(u,v;\boldsymbol{x_p})|^2 = \left|\sum_m\sum_n \exp\left\{\frac{-jk}{2f_2}[(mT_2)^2+(nT_2)^2]\right\} \times g_2^{mn}\left(\frac{u}{\lambda l_2}, \frac{v}{\lambda l_2}; \boldsymbol{x_p}\right)\right|^2. \quad (5)$$

In the derivations, all lenses are assumed to be thin and aberration-free. For more details of the PSF derivation for plenoptic cameras, we refer the reader to [8].

Having specified the PSF, the discrete sensor data at the output of the plenoptic camera is approximated as

$$I[i,j] \approx \sum_{m=1}^{N}\sum_{n=1}^{N} I_h[m,n]\mathrm{PSF}[i,j;\boldsymbol{x}^{mn}], \quad (6)$$

where $\boldsymbol{x}^{mn}$ denotes the sampling points on the approximating surface (marked by circles in Fig. 1), $\mathrm{PSF}[i,j;\boldsymbol{x}^{mn}]$ are the corresponding discrete PSFs, and $I_h[m,n]$ represents the desired high-resolution image of the scene. Please note that, in the simulations, the discretization on the sensor plane is implemented by sampling the continuous signals on an oversampled sensor grid and then summing up the sampled values inside the pixels. If we vectorize the matrices $I_h[m,n]$ and $I[i,j]$ into $N^2 \times 1$ vectors $\boldsymbol{x}$ and $\boldsymbol{b}$, respectively, and form the imaging system matrix $\boldsymbol{P}$ from PSFs according to the vectorization used for $\boldsymbol{x}$ and $\boldsymbol{b}$, the super-resolution problem can then be formulated as estimation of $\boldsymbol{x}$ given that

$$\boldsymbol{b} = \boldsymbol{P}\boldsymbol{x} + \boldsymbol{n}, \quad (7)$$

where $\boldsymbol{n}$ represents the additive noise.

For the typical assumption of independent and identically distributed (i.i.d.) Gaussian noise, a general class of solutions

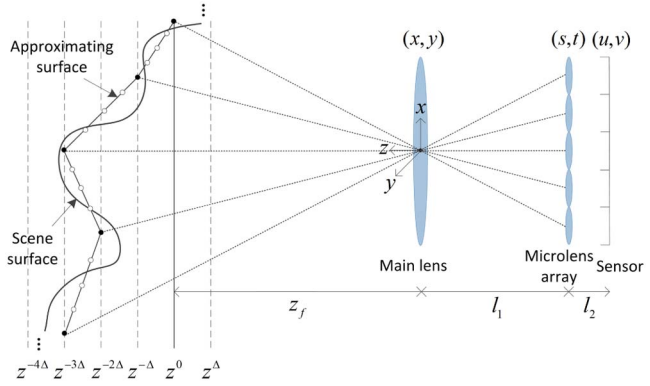

Fig. 1. Defocused plenoptic imaging system and approximation of the scene (not drawn to scale).

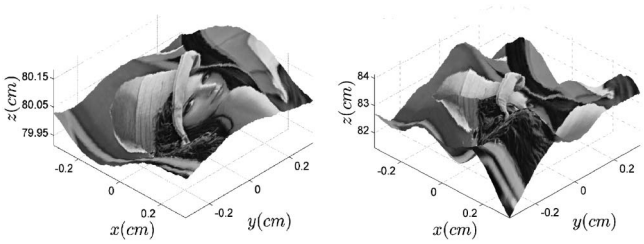

Fig. 2. Curved surfaces centered at (a) $z_c = 800.5$ mm and (b) $z_c = 828$ mm.

for the problem given in Eq. (7) can be obtained by the following regularized least squares minimization:

$$\hat{\boldsymbol{x}} = \min_{\boldsymbol{x}} \|\boldsymbol{P}\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \alpha\Phi(\boldsymbol{x}), \tag{8}$$

where $\Phi(\boldsymbol{x})$ is the regularization function, and α is the regularization coefficient. As a consequence of the highly ill-posed nature of the super-resolution problem for in-focus scene regions, the choice of the regularization function is of critical importance to get a satisfactory solution. The sparse image derivative priors have been shown to effectively model natural images and thus serve as an adequate regularization in many image restoration problems [14,15]. In this Letter, we utilize the prior $\Phi(\boldsymbol{x}) = |\nabla_x \boldsymbol{x}|^{0.8} + |\nabla_y \boldsymbol{x}|^{0.8}$, where ∇_x and ∇_y are the first-order image derivative operators. We employ the iterative re-weighted least squares method to solve the resulting non-convex optimization problem [14].

The developed algorithm is tested for in-focus and near in-focus regions by several simulations performed for various 2D surfaces modeling the scene. The common plenoptic camera parameters for all simulations are set as $f_1 = 35$ mm, $T_1 = 18.4$ mm, $f_2 = 25$ μm, $T_2 = 13.89$ μm, and the sensor pixel pitch is set to 1.4 μm. These parameters are chosen in accordance with the commercially available defocused plenoptic camera [16]. The microlens array constitutes of square-shaped microlenses which are placed on an orthogonal grid with a fill-factor 90%. Other simulation parameters are chosen as $\lambda = 534$ nm, $z_f = 800$ mm, $M = 21$, and $N = 210$.

The curved surfaces are centered at $z_c = 800.5$ mm and $z_c = 828$ mm for the in-focus and near in-focus scene cases, respectively. For a point source at $z = 828$ mm, the diameter of the blur spot produced on the microlens array plane is twice the microlens size. Thus, the corresponding PSF spreads over two elemental images. For the in-focus scene case, the surfaces are kept smoother but are still representative as part of a 3D scene, so that all points on it are as close as possible to the focused plane at 800 mm. The surfaces used in the simulations are shown in Fig. 2.

The super-resolution solution given by Eq. (8) is obtained for several noise (i.i.d. Gaussian) levels and low-resolution depth maps with different accuracies (determined by the parameter Δ). The depth maps are actually derived from disparity maps which are specified for the center view with respect to adjacent sub-aperture views along the horizontal and vertical directions. In addition to finite disparity steps that correspond to finite depth accuracies, two ideal cases are also included: continuous, but still low resolution, depth which corresponds to an infinitesimal disparity step (denoted by ϵ), and continuous high-resolution depth which gives the exact (ground-truth) information about the high-resolution texture sample positions on the curved surface. For the cases where a low-resolution depth map is utilized, the high-resolution depth maps are approximated using bi-linear interpolation. The reconstruction peak signal-to-noise ratios (PSNRs), obtained for the curved surfaces shown in Fig. 2, are presented in Table 1. In PSNR calculation, the ideal high-resolution image is taken to be the one that would be formed on the microlens array plane by a pinhole camera at the center of the main lens.

Examples of reconstructed images are shown in Figs. 3(b) and 3(c). The ray optics model is also implemented, using geometric optics principles as in [6], for comparison purposes. The resulting reconstructed images, for the same parameters as in Fig. 3(b), are shown in Fig. 3(a).

Table 1 and Fig. 3 reveal important observations. Regarding the comparison between the ray and wave optics models, as illustrated in Figs. 3(a) and 3(b), the wave optics model provides considerably higher quality reconstructions for both

Table 1. Reconstruction PSNRs

(a) PSNRs for the Surface Centered at $z_c = 800.5$ mm; Means and Standard Deviations (μ,σ) Are Calculated over 30 Runs

				Δ (pixels)							
		Ground-Truth		ϵ		0.005		0.01		0.05	
PSNR (dB)		μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
SNR	60 dB	25.96	0.21	25.74	0.06	22.03	0.04	20.95	0.01	19.25	0.03
	50 dB	24.29	0.05	24.23	0.13	21.95	0.18	20.91	0.06	19.19	0.08
	40 dB	22.48	0.12	22.47	0.16	21.44	0.31	20.56	0.21	18.83	0.18

(b) PSNRs for the Surface Centered at $z_c = 828$ mm; Means and Standard Deviations (μ,σ) Are Calculated over 30 Runs

				Δ (pixels)							
		Ground-Truth		ϵ		0.01		0.05		0.1	
PSNR (dB)		μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
SNR	50 dB	26.82	0.40	25.36	0.09	25.02	0.05	22.56	0.01	22.00	0.01
	40 dB	25.10	0.43	24.90	0.12	24.62	0.24	22.55	0.03	21.98	0.01
	30 dB	23.70	0.11	23.66	0.11	23.61	0.17	22.42	0.04	21.89	0.04

Fig. 3. Super-resolution results for SNR = 40 dB. (a) Ray optics model. (b), (c) Wave optics model. Top: $z_c = 800.5$ mm; (a), (b) $\Delta = 0.005$, (c) $\Delta = 0.01$. Bottom: $z_c = 828$ mm; (a), (b) $\Delta = 0.01$, (c) $\Delta = 0.05$.

in-focus and near in-focus regions. Due to highly ill-posed nature of the problem, especially for $z_c = 800.5$ mm, the reconstructed image quality strongly depends on the level of depth discretization error. For example, a sub-pixel accurate disparity map with step size $\Delta = 0.005$, which corresponds to a depth discretization step of ~0.7 mm around $z = 800.5$ mm, yields a relatively good reconstruction [top row of Fig. 3(b)]. When the step size is increased by a factor of 2, apparent degradations become visible in the reconstructed image [as marked on the image on the top row of Fig. 3(c)]. The simulations, on the other hand, also demonstrate that if there exist only few pixels of blur in sub-aperture images, as in the case of $z_c = 828$ mm, the solution is observed to be significantly more robust to both depth errors and sensor noises.

In our experiments we have assumed a rather simple depth upsampling method. As Table 1 shows, for the two ideal cases (ground-truth and ϵ columns), it causes acceptable reconstruction errors. For more complex scenes, the result will also depend on how well the depth upsampling method handles complexities such as occlusions. Thus, one should employ more sophisticated algorithms, e.g., as in [17], to tackle such cases.

In conclusion, we have demonstrated super-resolved images of in-focus scene regions for the defocused plenoptic camera. The depth accuracy analysis has indicated that rather high accuracies are required to have reasonable reconstruction qualities for those regions. Nevertheless, we have shown that few pixels of blur in sub-aperture images (the case for near in-focus regions) can make super-resolution considerably more robust to both image noise and depth errors. On the other hand, the presented results have also demonstrated that by employing the wave optics model, significant improvements can be achieved, in super-resolution of both in-focus and near in-focus scene cases, compared to the ray optics model. We address exploring more robust super-resolution schemes as the critical aspect of the problem for in-focus scene regions, in order to have improved reconstruction quality or a relieved accuracy requirement on the depth map.

Funding. European Commission (EC) (32449); Suomen Akatemia (Academy of Finland) (137012, 287150).

REFERENCES

1. G. Lippmann, J. Phys. **7**, 821 (1908).
2. E. H. Adelson and J. Y. A. Wang, IEEE Trans. Pattern Anal. Mach. Intell. **14**, 99 (1992).
3. R. Ng, "Digital light field photography," Ph.D. thesis (Stanford University, 2006).
4. A. Lumsdaine and T. Georgiev, in *IEEE International Conference on Computational Photography (ICCP)* (IEEE, 2009), pp. 1–8.
5. S. C. Park, M. K. Park, and M. G. Kang, IEEE Signal Process. Mag. **20**, 21 (2003).
6. T. E. Bishop and P. Favaro, IEEE Trans. Pattern Anal. Mach. Intell. **34**, 972 (2012).
7. C. K. Liang and R. Ramamoorthi, ACM Trans. Graph. **34**, 1 (2015).
8. S. A. Shroff and K. Berkner, Appl. Opt. **52**, D22 (2013).
9. I. Tosic, S. A. Shroff, and K. Berkner, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2013), pp. 1821–1825.
10. P. Helin, V. Katkovnik, A. Gotchev, and J. Astola, in *3DTV-Conference: The True Vision—Capture, Transmission and Display of 3D Video (3DTV-CON)* (IEEE, 2014), pp. 1–4.
11. M. Broxton, L. Grosenick, S. Yang, N. Cohen, A. Andalman, K. Deisseroth, and M. Levoy, Opt. Express **21**, 25418 (2013).
12. H. G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y. W. Tai, and I. S. Kweon, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2015), pp. 1547–1555.
13. J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, 1996).
14. A. Levin, R. Fergus, F. Durand, and W. T. Freeman, ACM Trans. Graph. **26**, 70 (2007).
15. T. Georgiev and A. Lumsdaine, "Superresolution with plenoptic camera 2.0," Technical Report (Adobe Systems, 2009).
16. "The first generation lytro camera," https://store.lytro.com/collections/the-first-generation-lytro-camera.
17. M. Hornacek, C. Rhemann, M. Gelautz, and C. Rother, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2013), pp. 1123–1130.