DECONVOLUTION BASED LIGHT FIELD EXTRACTION FROM A SINGLE IMAGE CAPTURE

M. Zeshan Alam and Bahadir K. Gunturk

Dept. of Electrical and Electronics Engineering, Istanbul Medipol University, Istanbul, Turkey

ABSTRACT

In this paper, we propose a method to extract light field using a conventional camera from a single image capture. The method involves an offline calibration process, where point spread functions, relating different perspective images captured with a narrow aperture to a central image captured with a wide aperture, are estimated for different depths. During application, light field perspective images are recovered by deconvolving the input image with the set of point spread functions that were estimated in the offline calibration process.

Index Terms— light field imaging, depth estimation from a single image

1. INTRODUCTION

With a light field (also known as plenoptic) imaging device, the intensities of the light rays in different directions are recorded separately at each pixel position. The addition of the angular information enables new post-capture capabilities, such as digital refocusing, aperture size control, perspective shift, and depth estimation [1, 2].

Light field acquisition can be done in a variety of ways including micro-lens arrays (MLAs) [3, 2, 4], coded masks [5], [6], camera arrays [7, 8], and camera attached to a gantry [9]. Among these different implementations, MLA based light field cameras offer a cost-effective approach, leading to commercial light field cameras [10, 11]. The main drawback of the MLA based light field cameras is low spatial resolution since a single sensor is shared to record both angular and spatial information. To overcome the spatio-temporal resolution trade-off, several methods have been developed, including improved interpolation of light field data [12], super-resolution reconstruction [13], convolutional neural network based super-resolution reconstruction [14], and hybrid cameras [15, 16, 17].

Coded mask based light field cameras offer an alternative to MLA based light field cameras. There are various design approaches, including multiple-capture [18, 5, 6, 19], multiple-mask [20] and single-capture [21] methods. While the multiple-capture methods are limited to applications with static scenes, single-capture methods may suffer from low light efficiency. Since a single sensor is used to obtain both angular and spatial information, these cameras may also have the low spatial resolution problem, as in the case of MLA based cameras.

Camera arrays and gantry based designs do not have a spatial resolution issue, but they are bulky and have less mobility. The angular resolution in camera arrays is limited by how close the cameras can be placed. Gantry based systems do not have such an angular resolution limitation, however, they are not suitable for dynamic scenes.

In this paper, we present a light field acquisition method using a regular camera from a single capture. The method involves an offline calibration process, during which point spread functions (PSFs), relating images captured with a narrow aperture from different perspectives to a central image captured with a wide aperture, are estimated for different depths. The set of PSFs for different perspectives and depths is saved for the light field extraction process. To extract the light field for a scene, the aperture of the camera is set wide as in the case of the central image in the offline process, and a single image is taken. Using the PSFs obtained during the offline calibration process, individual perspective images are going to be recovered. Since there is no depth information about the scene, the depth of each pixel should be estimated as well. For a specific perspective, the input image is deconvolved with the corresponding PSF perspective and for each PSF depth. The correct PSF scale (i.e., depth) for a pixel is determined based on an energy function that returns small value when the correct scale is used. The selective fusion of the deconvolved image regions based on the depth labels forms the perspective image. The process is repeated for all perspectives to recover the entire light field.

The proposed method overcomes the spatio-angular trade-off and enables high-resolution light field capture from a single capture. In our experiments, we obtained light fields with 11 x 11 angular resolution and 1024 x 1280 spatial resolution, which is the resolution of the sensor in the camera that we used. We compared our results with light field captured by a first-generation Lytro camera, and demonstrated the resolution improvement. In addition to the resolution issue of the MLA based light field cameras, the proposed method also overcomes the poor light efficiency issue of the coded mask based light field cameras. On the downside, the proposed

The work is partially supported by Istanbul Development Agency.

method involves deconvolution, which may introduce artifacts.

2. OFFLINE PSF ESTIMATION FOR DIFFERENT PERSPECTIVES AND DEPTHS

The offline calibration process is illustrated in Figure 1. A planar object is placed at a distance d_j from the camera. The aperture of the camera is closed to a narrow opening and the camera is moved to a perspective position p_i. The image corresponding to that specific perspective and depth is then captured. We denote this image as $I_{i,j}(x, y)$, where the subscripts (i, j) indicates the perspective position and the object depth, and (x, y) indicates a pixel coordinate. In addition to the narrow aperture perspective images, an image with a wide aperture, denoted as $B_j(x, y)$, is also taken for the same depth. The process is repeated for all depths and perspectives. The perspective positions are chosen on a regular grid within the wide aperture opening.

The wide aperture image $B_j(x, y)$ can be modeled as the superposition of narrow aperture images taken from different perspectives within the aperture. Neglecting the occluded regions, the image $B_j(x, y)$ can be written as the convolution of the narrow aperture image $I_{i,j}(x, y)$ with a PSF $k_{i,j}(x, y)$: $B_j(x, y) = k_{i,j}(x, y) * I_{i,j}(x, y)$, which would result in a set of linear equation when written for all (x, y). We can write these equations as $\mathbf{B}_j = \mathbf{I}_{i,j}\mathbf{k}_{i,j}$, where $\mathbf{B}_j$ and $\mathbf{k}_{i,j}$ are the vectorized forms of the wide aperture image and the PSF, and $\mathbf{I}_{i,j}$ matrix is constructed from $I_{i,j}(x, y)$. We solve this system using the least squares estimation technique to obtain the PSF for a specific perspective and depth. The PSF estimation process is repeated for all depths and perspectives.

Our prototype system is shown in Figure 2; it includes two motorized translation stages (Thorlabs NRT150) and a regular camera with a 1024 x 1280 CMOS sensor and a 35mm lens. During the calibration, the planar target object is moved within a depth range of 2 meters from the camera with steps of 10cm. For each depth, the camera is moved with a step size of 0.1mm as shown in the figure to capture the perspective images; in addition, the camera is moved to the central position to capture the wide aperture image.

3. DECONVOLUTION BASED LIGHT FIELD EXTRACTION

Light field extraction from a single wide aperture image capture consists of several steps. First, for each perspective, a pixel-wise depth (PSF scale) map is obtained. The depth map is then utilized to fuse multiple deconvolved images to construct a perspective image. The process is repeated for all perspectives.

Our PSF scale identification algorithm is adopted from [22], which exploits the reconstruction error when there is scale and blur mismatch. When an image of a natural scene

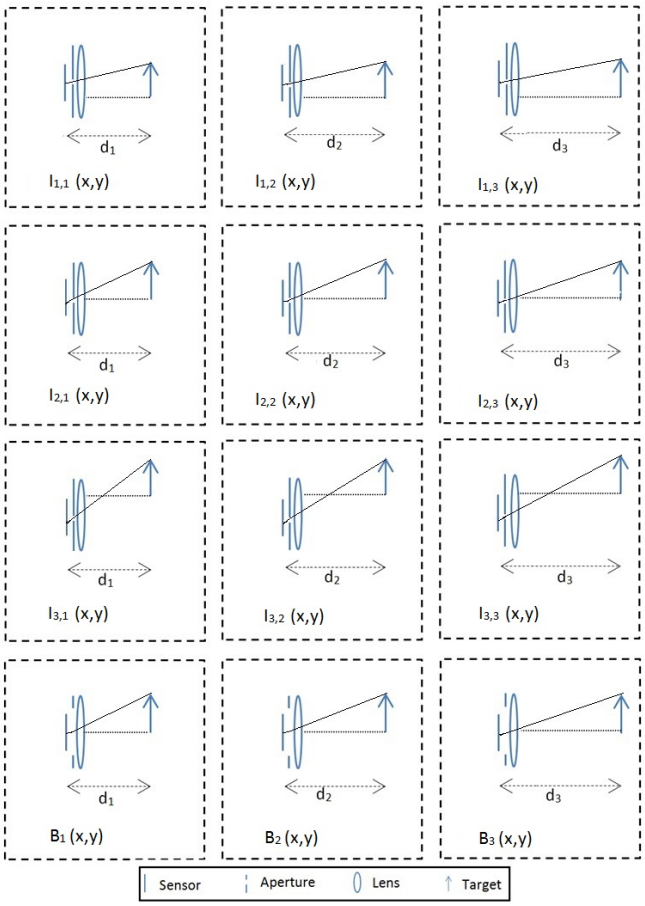

Fig. 1: Illustration of the offline calibration process. For each depth, narrow aperture images from different perspective locations and a single wide aperture image are taken. The process is repeated for different depths.

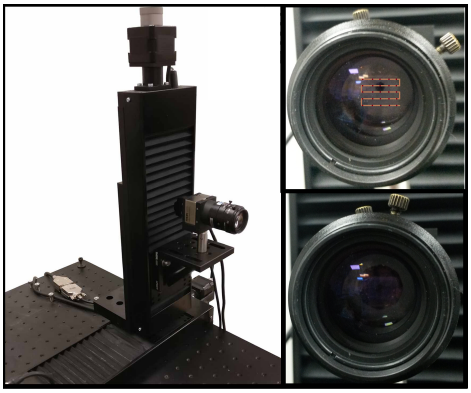

Fig. 2: Setup for the offline calibration process. (Left) Motorized translation stages and the camera. (Right Top) Narrow aperture opening for perspective images. The camera is moved as illustrated on lens for different perspective images. (Right Bottom) Wide aperture opening for the central image.

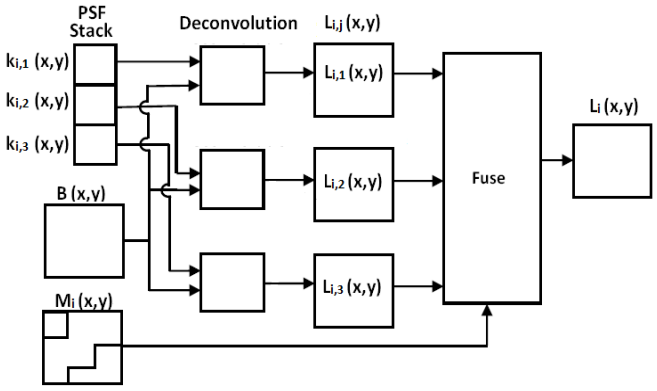

Fig. 3: Single perspective image recovery process from an image.

with multiple depths is deconvolved with the PSF of a particular scale, the corresponding depth regions in the image become sharp while severe ringing artifacts appear in the rest of the image because these areas cannot be explained by that scale of the PSF. These ringings are dense and have gradients with magnitude significantly larger than that of natural sparsely distributed data; the difference can be used as a cue for kernel scale identification [22, 23]. We used the criterion in [22] to determine the correct PSF scale for each pixel of the input image.

The resulting scale (depth) map $M_i(x, y)$ may have some misclassified labels or gaps in the depth map. This may be due to occluded areas or regions without sufficient texture. While inpainting techniques can be used for view synthesis [24, 25], we preferred a simple approach in this work. We applied a mode filter, that is, a sliding window over the depth map to compute the mode of a local neighborhood and re-assign the center pixel to the mode label.

The recovery of a perspective image using the depth map is illustrated in Figure 3. First, the input image $B(x, y)$ is deconvolved for every color channel using [22] with the PSFs, known through the previous PSF scale identification step. The deconvolved image $L_{i,j}(x, y)$ has one of the depths recovered while ringing artifacts appear on regions other than the recovered depth. The depth map $M_i(x, y)$ is used to generate a set of binary masks corresponding to the PSF labels; and these masks applied on corresponding deconvolved images to construct a single perspective image $L_i(x, y)$. The process is repeated for all perspectives.

4. EXPERIMENTAL RESULTS

The proposed method is tested on real data and compared with a first generation Lytro camera. In Figure 4, we show a subset of the PSFs obtained in the offline calibration process. In Figure 5, we present an input image $B(x, y)$, images deconvolved with the PSFs corresponding to the middle perspective and three different depths, the label map and the final perspective image. From the deconvolved images, it is seen that when the PSF scale does not match the region, ringing artifacts occur. The final image is constructed from the first two deconvolved images given in the figure.

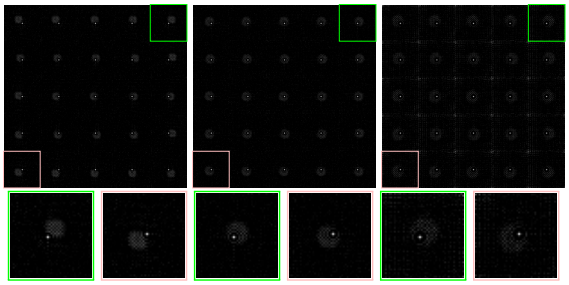

Fig. 4: The set of PSFs shown for 5 x 5 perspectives and three different depths.

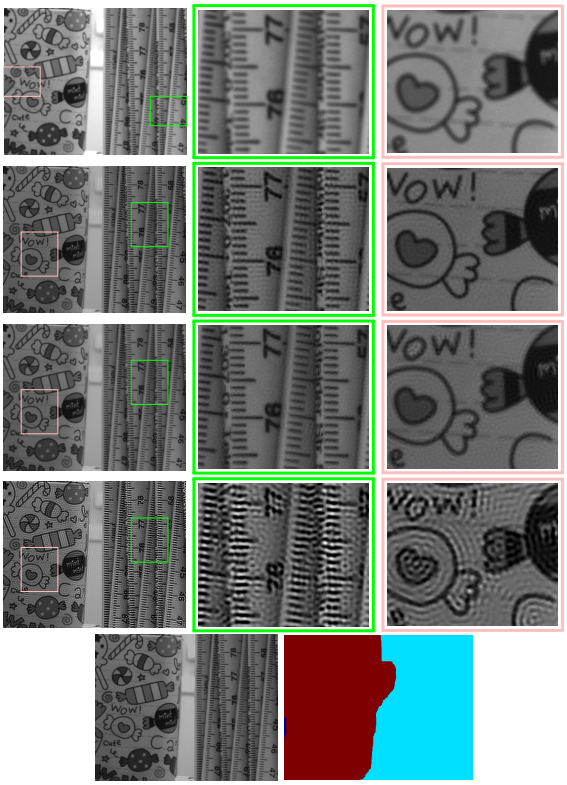

Fig. 5: Recovery of a perspective image. (Row 1) Input image. (Row 2-4) Deconvolved images with different PSFs. The first two deconvolved images have matching depths, which can be identified from the zoomed-in regions; the last deconvolved image does not have any matching depth. (Row 5) Recovered perspective image and the label map indicating the regions taken from the first two deconvolved images. The PSFs used in the deconvolution are given in Figure 4.

Fig. 6: Comparison of light field perspective image recovered by the proposed method and captured by a Lytro camera. Input image is the wide aperture image used by the proposed method.

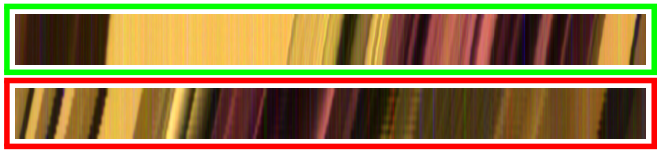

Fig. 7: Epipolar plane images of the 11 x 11 reconstructed light field. (Top) Horizontal EPI. (Bottom) Vertical EPI.

In Figure 6, we compare a perspective image obtained with our method and a perspective image produced by a Lytro camera. The proposed method produces a sharper perspective image than the Lytro camera does. In Figure 7, we present horizontal and vertical epipolar plane images (EPIs) of the light field generated using the proposed method.

In Figure 8, we perform light field refocusing, achieved using the shift and sum technique, on an 11 x 11 light field generated using the proposed method and a light field captured by a Lytro camera. It is seen that the proposed method produces higher resolution results.

5. CONCLUSIONS

In this paper, we propose a method to recover light field from a single capture with a regular camera. The method is based on deconvolving a wide aperture image with a set of PSFs calculated in an offline calibration process. It has good light efficiency and can achieve spatial resolution as much as the camera sensor has since it does not have any spatio-angular tradeoff unlike the MLA based light field cameras. The method can be used to convert any regular camera with a controllable aperture into a light field camera.

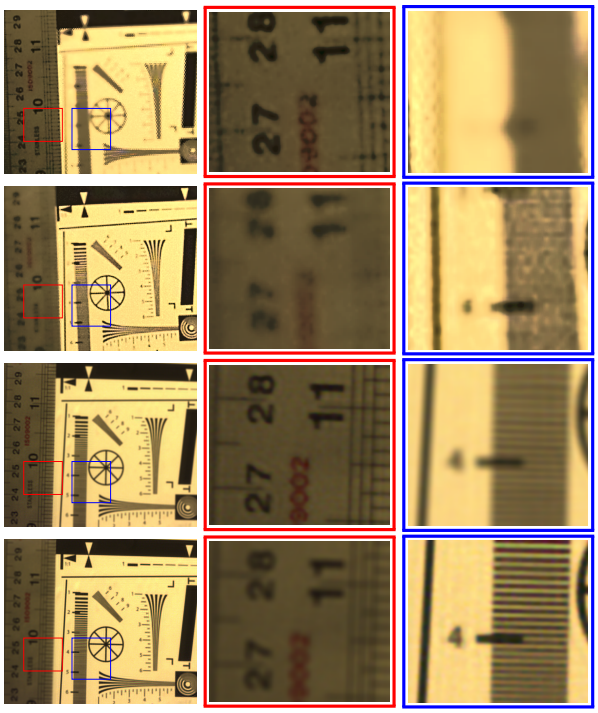

Fig. 8: Comparison of post-capture refocusing. (Row 1) Lytro close focus. (Row 2) Lytro far focus. (Row 3) Proposed method close focus. (Row 4) Proposed method far focus.

6. REFERENCES

[1] M. Levoy and P. Hanrahan, "Light field rendering," in *ACM Int. Conf. on Computer Graphics and Interactive Techniques*, 1996, pp. 31–42.

[2] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," in *Stanford University Computer Science Technical Report CSTR*, 2005.

[3] E. H. Adelson and J. Y. Wang, "Single lens stereo with a plenoptic camera," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 99–106, 1992.

[4] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *IEEE Int. Conf. on Computational Photography*, 2009, pp. 1–8.

[5] A. Ashok and M. A. Neifeld, "Compressive light field imaging," in *SPIE Defense, Security, and Sensing*, vol. 7690, 2010.

[6] S. D. Babacan, R. Ansorge, M. Luessi, P. R. Mataran, R. Molina, and A. K. Katsaggelos, "Compressive light field sensing," *IEEE Trans. on Image Processing*, vol. 21, pp. 4746 – 4757, 2012.

[7] B. Wilburn, N. Joshi, V. Vaish, E. V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Trans. on Graphics*, vol. 24, pp. 765–776, 2005.

[8] J. C. Yang, M. Everett, C. Buehler, and L. McMillan, "A real-time distributed light field camera," in *Eurographics Workshop on Rendering*, 2002, pp. 77–86.

[9] J. Unger, A. Wenger, T. Hawkins, A. Gardner, and P. Debevec, "Capturing and rendering with incident light fields," in *Eurographics Workshop on Rendering*, 2003, pp. 141–149.

[10] "Lytro, Inc." https://support.lytro.com/hc/en-us/, accessed: 2018-05-24.

[11] "Raytrix, gmbh," https://www.raytrix.de/, accessed: 2018-05-24.

[12] 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 *IEEE Int. Conf. on Computer Vision*, 2013, pp. 3280–3287.

[13] S. Wanner and B. Goldluecke, "Spatial and angular variational super-resolution of 4D light fields," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 2012, pp. 901–908.

[14] M. S. K. Gul and B. K. Gunturk, "Spatial and angular resolution enhancement of light fields using convolutional neural networks," *IEEE Trans. on Image Processing*, vol. 27, pp. 2146–2159, 2018.

[15] V. Boominathan, K. Mitra, and A. Veeraraghavan, "Improving resolution and depth-of-field of light field cameras using a hybrid imaging system," in *IEEE Int. Conf. on Computational Photography*, 2014, pp. 1–10.

[16] M. Z. Alam and B. K. Gunturk, "Hybrid light field imaging for improved spatial resolution and depth range," *Machine Vision and Applications*, vol. 29, pp. 11–22, 2018.

[17] M. U. Mukati and B. K. Gunturk, "Hybrid-sensor high-resolution light field imaging," in *25th Signal Processing and Communications Applications Conf.*, 2017, pp. 1–4.

[18] C. K. Liang, T. H. Lin, B. Y. Wong, C. Liu, and H. H. Chen, "Programmable aperture photography: multiplexed light field acquisition," *ACM Trans. on Graphics*, vol. 27, pp. 55:1–55:10, 2008.

[19] Y. P. Wang, L. C. Wang, D. H. Kong, and B. C. Yin, "High-resolution light field capture with coded aperture," *IEEE Trans. on Image Processing*, vol. 24, pp. 5609 – 5618, 2015.

[20] Z. Xu and E. Y. Lam, "A high-resolution lightfield camera with dual-mask design," in *SPIE Optical Engineering+ Applications*, vol. 8500, 2012.

[21] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, "Compressive light field photography using overcomplete dictionaries and optimized projections," *ACM Trans. on Graphics*, vol. 32, pp. 46:1–46:12, 2013.

[22] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. on Graphics*, vol. 26, 2007.

[23] J. Lin, X. Lin, X. Ji, and Q. Dai, "Separable coded aperture for depth from a single image," *IEEE Signal Processing Letters*, vol. 21, pp. 1471–1475, 2014.

[24] I. Daribo and B. Pesquet-Popescu, "Depth-aided image inpainting for novel view synthesis," in *IEEE Int. Workshop on Multimedia Signal Processing*, 2010, pp. 167–170.

[25] K. Oh, S. Yea, and Y. Ho, "Hole filling method using depth based in-painting for view synthesis in free viewpoint television and 3D video," in *IEEE Picture Coding Symposium*, 2009, pp. 1–4.