라이트 필드 영상의 공간해상도 개선: 리뷰

*임종훈, Vinh Van Duong, 전병우
성균관대학교 전자전기컴퓨터공학과
e-mail: {*yjh6522, duongvinh, bjeon*}*@skku.edu*

Light Field Image Spatial Resolution Enhancement: A Review

JonghoonYim, Vinh Van Duong, Byeungwoo Jeon
Department of Electrical and Computer Engineering
Sungkyunkwan University

Abstract

Light Field (LF) cameras capture both spatial and directional information of light rays. Current LF cameras have a problem of a low spatial resolution. There have been lots of existing works carried out to improve the resolution of LF images. In this paper, those existing works will be divided into two categories: hardware-based approaches and software-based approaches, and we will look into and compare several experiment results in order for LF spatial resolution enhancement. Finally, the direction for the future spatial resolution enhancement will be suggested.

1. Introduction

Unlike conventional cameras that record irradiance on each pixel, LF cameras also capture the direction of light ray coming through the lens. LF imaging thus can provide lots of new functionalities such as post-capture refocusing, viewpoint change, 3D modeling, and depth estimation [1, 2, 5]. It becomes possible due to its unique structure: Micro Lens Array (MLA), which is located between the main lens and the image sensor.

Today's LF cameras are portable and commercially available, allowing us to use those functionalities easily. Moreover, it is being utilized in the fields of biological microscopy, background and object extraction, detecting defects in a machine, and 3D analysis of a plant growth [5, 15].

However, current LF cameras suffer from low spatial resolution because a single image sensor is used to record both spatial and angular data of a light field. In the first generation, the Lytro camera has one 11 megapixel-imaging sensor, but it can produce an image with a spatial resolution of 0.1 megapixels, which is about 1/100 of the number of sensor pixels [5]. Therefore, in order to acquire an image with better quality using current LF technology, spatial resolution enhancement is necessary. Low resolution also causes a problem in refocusing application because it brings unwanted artifacts on the image after processing [1, 21].

There have been lots of works done to enhance the spatial resolution of the images from LF cameras. In this paper, they are divided into two groups: hardware-based approaches, and software-based approaches, and reviewed in Sections 2 and 3, respectively, concluding remarks will be given in Section 4.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2020R1A2C2007673).

2. Hardware-based approaches

In this section, we will look into experiments for spatial resolution enhancement of light field image based on hardware. They tried to capture extra real data physically. We categorize them further into micro scanning [4, 5] and hybrid imaging [6, 7, 8].

2.1. Micro Scanning

There were researches to improve the spatial resolution by modifying the structure of the imaging device. In [4], 3D images were taken using a conventional camera and MLA. MLA is located in front of a camera, and is finely shifted by a 2D translational stage by an amount of half or one fourth of pixel pitch. After synthesizing the images from each capture, integral Fourier hologram system (IFHS) process was applied to create a 3D image. This method removed the aliasing (i.e., repetition of an object) from the image, as can be seen in Figure 1.

In [5], on the other hand, image sensor of the first generation Lytro light field camera was finely shifted by an additional equipment and software. In order to sample the light field, for example, twice or four times more densely, the sensor was shifted by the amount of half, or one fourth of the distance between the centers of two adjacent micro-lens. Spatial resolution was 4×4 times improved in [5] by synthesizing 16 light field captures of a static scene. It also achieved about 3 times higher maximum spatial frequency than that of an image from one light field capture.

2.2. Hybrid Imaging

Some experiments enhanced the spatial resolution of the light field images by taking pictures with a light field camera and a conventional DSLR camera [6, 7, 8]. Among them, in [6], a Lytro camera of 380×380 resolution together with one 18 megapixel Canon T3i DSLR camera took the same static scene (but not exactly at the same location). Following, a patch matching

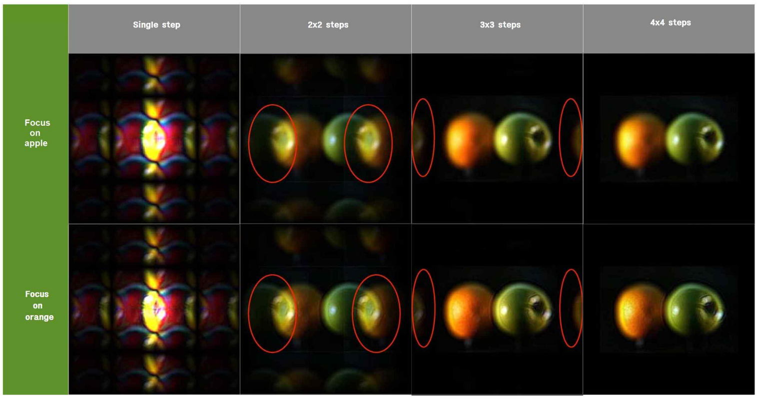

Figure 1: 3D images created by the method proposed in [4]. Images are generated with one step, 2*x*2, 3x3, and 4x4 steps micro-scanning. As steps get higher, the alias of the orange and apple disappeared [4].

algorithm was applied to form a high resolution multiview-image (part from each sub aperture image of LF is matched with the most correlating part of high-resolution image, which is found using an algorithm in [22]). It showed that the prediction error and prediction uncertainty [23] of super-resolved images are significantly lower compared to the method in [24]. M. Zhao, et al. [8] employed a cross scale resolution approach (CSR) instead of the patch matching algorithm in enhancing spatial resolution LF images. Compared to the method in [6], CSR preservedhigh-frequency information better, achieved lower RMSE in depth estimation.

3. Software-based approaches

We look into and compare the super resolution researches that did not rely on additional equipment for taking light field images. We use the same classification as [2]: projection-based [9]-[11], optimization-based [12]-[14], and learning-based [15]-[18]. The projection-based methods utilize how light field cameras capture the image. Lim et al. [11] used pixels from other sub aperture images (which is sub-pixel shift in spatial domain) to enhance the resolution. The optimization-based methods use a variety of optimization frameworks and mathematical models. Bishop et al. [13] used a Bayesian deconvolution approach on depth information from a light field. Learning-based model utilizes deep learning techniques to perform mapping of low-resolution light field into higher ones. Y. Yoon, et al. [15] employed two cascaded Convolutional Neural Networks (CNN) to enhance both spatial and angular resolution of LF images.

In order to compare the various methods, Cheng et al. [2] performed spatial resolution enhancement of the light field images with various methods and applied the same metrics for evaluation. Considering the quality difference and different size of test data, light field images were used from two datasets: HCI synthetic data set [19] and EPFL real-world dataset [20]. Also, two degradation kernels were used: Bicubic, Gaussian, with downscaling factors of 2 and 3. Then, bicubic interpolation (BIC) and four other methods from three classifications were selected for the evaluation: projection-based method (PRO) [10], optimization-based using graphs (GB) [14], and two from learning based methods: ridge regression (RR) [17], and CNN-based method (LFCNN) [16]. After spatial resolution enhancement of the light field images, in order to evaluate the results, PSNR, SSIM, and VGG value [25] (a distance between the original image and super-resolved image in feature space) were used as comparison criteria. Here, PRO generates high-resolution of only sub aperture image (SAI) of the central view. Therefore, evaluation was conducted for other methods by averaging the results of super-resolved SAIs.

Table 1 lists average PSNR (the higher, the better) and VGG values (the lower, the better) for the five methods over two datasets and four degradation kernels. All methods outperform BIC method for every case. It is shown that non-learning methods like PRO and GB gave better result than RR and LFCNN on HCI synthetic dataset, while the learning-based methods outperformed on EPFL real world dataset. The underlying reason is that the learning-based methods use neural networks trained with lots of real-world images, whereas PRO and GB, which employ the mathematical modeling of LF imaging are advantageous in synthetic datasets which have cleaner and simpler scene than that of EPFL dataset. LFCNN notably outperformed other methods for both kernels with downscaling factor of 2. For VGG value, which represents perceptual quality (the lower, the better), PRO excelled GB and RR, and slightly outperformed LFCNN for downscaling factor of 3. For visual results, PRO gave fine result in texture, while LFCNN gave clear borders and edges between the objects [26]. The results in terms of SSIM metric are given in [26]. Results of SSIM metric gave similar observations as those of PSNR results.

4. Discussions and Conclusion

In this paper, we have looked into several experiments for enhancing the spatial resolution of LF image. Experiments were divided into

Table 1. Performance Comparison (average PSNR [dB]/Average VGG values (with $\times 10\ e^4$)

Method			BIC	PRO	GB	RR	LFCNN
HCI Dataset	Bicubic	x2	36.28 / 0.2038	36.93 / 0.1276	37.25 / 0.1595	36.44 / 0.1459	**37.64** / **0.1071**
		x3	32.57 / 0.4242	33.84 / **0.3102**	**33.90** / 0.3476	33.61 / 0.3565	33.49 / 0.3352
	Gaussian	x2	34.28 / 0.3080	35.11 / 0.2188	35.12 / 0.2589	35.16 / 0.2303	**36.99** / **0.1314**
		x3	32.79 / 0.4045	**34.52** / **0.2407**	34.14 / 0.2956	33.49 / 0.3453	33.29 / 0.3206
EPFL Dataset	Bicubic	x2	31.45 / 0.2858	31.60 / 0.1936	31.88 / 0.2046	32.80 / 0.2194	**33.13** / **0.1373**
		x3	28.97 / 0.5894	29.21 / 0.4629	29.56 / 0.482	30.17 / 0.5047	**30.50** / **0.4244**
	Gaussian	x2	30.34 / 0.4315	30.59 / 0.3126	30.79 / 0.3517	31.68 / 0.3342	**33.25** / **0.1576**
		x3	28.97 / 0.5621	29.35 / **0.4112**	29.78 / 0.4203	29.96 / 0.5052	**30.11** / 0.4333

The lower VGG value means better performance. PSNR and VGG values are meansured from the centrla-view SAI after super-resolution enhancement using respectively five methods over datasets and degradation kernels [2]. The best and the second best are respectively **in bold** and underlined. The results for SSIM metric are shown in [26].

hardware-based methods and software-based methods. Hardware-based enhanced the spatial resolution by capturing additional true data using extra equipment, whereas software-based methods enhanced resolution using various mathematical models and software tools with already captured information. Since we have no common criterion for the performance comparison of the hardware-based works with the software-based ones, it is difficult to compare the performance directly. However, considering the additional equipment and setups, it is apparent that the methods based on hardware are not practical because we cannot bring and install all the instruments every time we take pictures. Therefore, research on light field spatial enhancement should be continued on the software-based methods, especially on the way that can combine the advantages of PRO and LFCNN.

References

[1] R. Ng, et al., "Digital Light Field Photography," *Ph.D. dissertation*, Stanford University, 2006.

[2] Z. Cheng, et al., "Light Field Super-Resolution: A Benchmark," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019.

[3] J. Jang, et al., "Improved viewing resolution of three-dimensional integral imaging by use of nonstationary micro-optics," in *OPTICS LETTERS*, vol. 27, no. 5, pp. 324-326, 2002.

[4] C. Yan, et al., "Reconstruction Improvement in Integral Fourier Holography by Micro-Scanning Method," in *Journal of Display Technology*, vol. 11, no. 9, pp. 709-714, 2015.

[5] M. U. Mukati, et al., "Light Field Super Resolution Through Controlled Micro-Shifts of Light Field Sensor," in *Signal Processing: Image Communication*, vol. 67, pp. 71-78, 2018.

[6] V. Boominathan, et al., "Improving Resolution and Depth-of-Field of Light Field Cameras Using a Hybrid Imaging System," in Proc. *IEEE Int. Conf. on Computational Photography (ICCP)*, 2014.

[7] J. Wu, et al., "A Novel Light Field Super-resolution Framework Based on Hybrid Imaging System," in Proc. *2015 Visual Communications and Image Processing (VCIP)*, 2015.

[8] M. Zhao, et al., "Cross-Scale Reference-Based Light Field Super-Resolution," in *IEEE Transactions on Computational Imaging*, vol. 4, no. 3, pp. 406-418, 2018.

[9] T. Georgiev, et al., "Superresolution with the focused plenoptic camera," in Proc. *SPIE Electronic Imaging 2011*, 2011.

[10] C.-K. Liang, et al., "A light transport framework for lenslet light field cameras," in ACM Transactions on Graphics (TOG), vol. 34, no. 2, p. 16, 2015.

[11] J. Lim, et al., "Improving the spatial resolution based on 4d light field data," in Proc. IEEE *ICIP*, pp. 1173-1176, 2009.

[12] S. Farag et al., "A novel disparity-assisted block matching-based approach for super-resolution of light field images," in Proc. *3DTV-Conference*, 2018.

[13] T. E. Bishop et al., "The light field camera: Extended depth of field, aliasing, and superresolution," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, 2012.

[14] M. Rossi, et al., "Light Field Super-Resolution Via Graph-based Regularization," in Proc. of *2017 IEEE 19th Int. Workshop on Multimedia Signal Processing (MMSP)*, 2017.

[15] Y. Yoon, et al., "Light Field Image Super-Resolution using Convolutional Neural Network," in *IEEE Signal Processing Letters*, vol. 24, no. 6, pp. 848-852, 2017.

[16] Y. Yoon, et al., "Learning a deep convolutional network for light-field image super-resolution," in

Proc. of *2015 IEEE Int. Conf. on Computer Vision Workshop (ICCVW)*, pp. 57-65, 2015.

[17] R. A. Farrugia, et al., "Super resolution of light field images using linear subspace projection of patch-volumes," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, 2017.

[18] Y. Wang, et al., "Lfnet: A novel bidirectional recurrent convolutional neural network for light-field image super-resolution," in *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4274-4286, 2018.

[19] S. Wanner, et al., "Datasets and benchmarks for densely sampled 4d light fields," in Proc. of *Annual Workshop on Vision, Modeling and Visualization*, pp. 225-226, 2013.

[20] M. Rerabek, et al., "New light field image dataset," in Proc. of *2016 Eighth Int. Conf. on Quality of Multimedia Experience (QoMEX)*, 2016.

[21] E. Lee et al., "Depth-based refocusing for reducing directional aliasing artifacts," in *Optics Express*, vol. 24, no. 24, 2016.

[22] M. Muja, et al., "Fast Approximate Nearest Neighbors With Automatic Algorithm Configuration," in Proc. of *VISAPP Int. Conf. on Computer Vision Theory and Applications*, vol. 1, pp. 331-340, 2009.

[23] M. Zontak, et al., "Internal Statistics of a Single Natural Image," in Proc. of *the IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 977-984, 2011.

[24] W.T. Freeman, et al., "Example-Based Super-Resolution," in *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56-65, 2002.

[25] R. Zhang et al., "The unreasonable effectiveness of deep features as a perceptual metric," in Proc. of *2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[26] Z. Cheng, et al., "Light Field Super-Resolution: A Benchmark Supplementary Material," in Proc. of the *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019.