Fast Channel Attention Network를 이용한 라이트필드 영상 복원 방법

아흐메드 살렘, 강현수*

충북대학교 정보통신공학부

ahmeddiefy@chungbuk.ac.kr, *hskang@cbnu.ac.kr

Light Field Image Reconstruction Using Fast Channel Attention Network

Ahmed Salem, Kang Hyun Soo*

Chungbuk National University

Abstract

Multi-view image reconstruction has been a hot topic recently. Multi-view images may increase many applications′ performance. This paper presents a technique for reconstructing high-density Light Field (LF) images. We apply our method to LF Lytro Illum images to address the trade-off between angular and spatial resolution induced by poor sensor resolution. In our system, view images are stacked and analyzed separately for speedy reconstruction. For LF reconstruction, we present a highly parallel residual channel attention network. Our technique leverages channel attention to find relationships between views and apply higher weights to relevant features, recovering more texture information. The suggested approach rapidly reconstructs high-quality images, according to experiments.

Keywords—Light field reconstruction; Angular resolution; Convolution neural network.

Ⅰ. Introduction

LF captures 3D scenes as dense, homogeneous images. These photos show 3D space and angles. Therefore, many applications have profited substantially from this vast quantity of information, like de-occlusion [1], and saliency detection [2]. By placing a microlens array in front of the image sensor, portable cameras may capture LF pictures [3]. Despite its benefits, this approach has poor sensor resolution, which makes it difficult to obtain LF images with high spatial and angular resolution. The LF reconstruction task builds dense LF images from sparse input views. Recently, learning-based techniques that improve LF reconstruction were proposed. They study the angular-spatial relationship but don′t fully employ the epipolar information [4,5].

This article discusses a learning-based approach for reconstructing densely sampled LF from poorly sampled LF. Our CNN model employs efficient convolutions to understand spatial-angular relationships. Our algorithm reconstructs a whole LF in a single forward pass after up-sampling it using Bicubic interpolation. We developed highly parallel residual network to rebuild 7 × 7 from 3 × 3 views. Our method aggregates view images into multi stacks. We use the channel attention for LF reconstruction [6]; with a highly parallel structure for quick reconstruction.

Ⅱ. Related Work

Wu et al. [4] suggested using "blur deblur" on epipolar images as a deep learning technique. Although this model performs better, it is more time-consuming and does not use light field data efficiently. Wang et al. [7] developed a pseudo-4D CNN for reconstructing LF with structure-preserving loss. They did, however, exacerbate error accumulation by hierarchically up-sampling LF. Liu et al. [8] used horizontal, vertical, and two-angle stacks to get more precise angle data than previous methods, but only one epipolar stack in each direction. Salem et al. [6] proposed using the channel attention to reconstruct LF images. Despite their network was fast, it was heavy.

Ⅲ. Methodology

LF images can be considered as a 2D array of view images with (H, W) and (U, V) spatial and angular resolution, as shown in Figure 1(a).

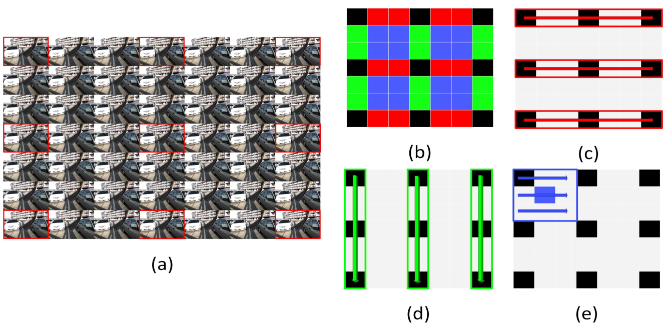

Figure 1. (a) 3 × 3 input red views to reconstruct 7 × 7 output views. (b) Input views shown in black, main views in red and green, and secondary views in blue. (c, d) Reconstructing stacks from main views located in rows and columns, respectively. (e) Reconstructing from secondary views.

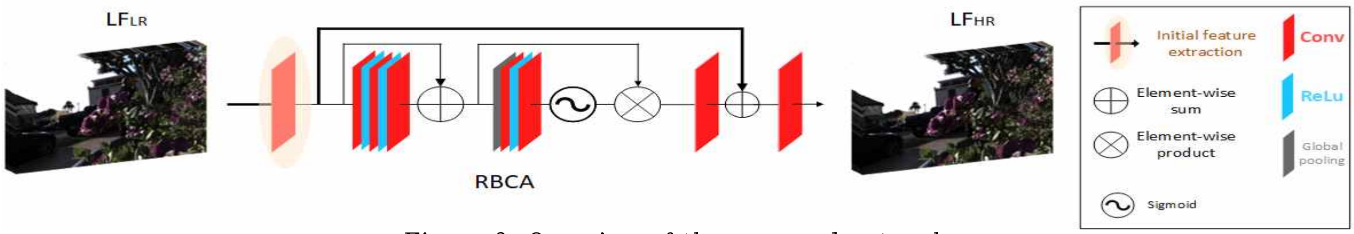

Figure 2. Overview of the proposed network.

Depending on the position of the view to be reconstructed, it can be considered a main or secondary view. The main views are located with the input in the same rows or columns. For main views, all the views are stacked to be fed to the network. We use the same CNN network to process rows and columns. Another CNN is used for the secondary views. Each secondary view is grouped with neighboring views to generate a stack to reconstruct this view, as shown in Figure 1 (b, c, d, e).

We proposed a very light model with three RBCA blocks " residual block cascaded by a channel attention block" that concentrate on the important features more, as shown in Figure 2. We used one convolution layer to extract initial features and to reconstruct the final image along with a long skip connection. This model is used to reconstruct the main and secondary views. The proposed network was trained to reduce the L1 distance between the raw LF input image and its corresponding ground-truth image.

We used the data provided in [12] for training and testing. Table 1 shows the numerical results where it's clear that our method outperforms other methods. Wu et al. [9] used epipolar images in a single direction; they up-sample LF, resulting in more mistakes in the final views. Yeung et al. [10]'s model generates spurious shadows and ghosting artifacts by ignoring the links between views. Liu et al. [8] used angular information more effectively than previous techniques; nonetheless, they used just one EPI stack per direction.

Table 2 illustrates the average run-time to reconstruct a full LF. Our model can reconstruct one LF in 1.27 seconds faster than other methods.

Table 1. Numerical comparison of the proposed model with the state-of-the-art models.

Dataset	Wu[9]	Ye[10]	Liu[8]	Our
30 Scenes	43.59/0.98	44.66/0.99	44.86/0.99	45.51/0.99
Reflective	43.09/0.97	43.9/0.98	44.31/0.98	44.84/0.98
Occlusion	39.75/0.95	40.0/0.95	40.16/0.96	41.3/0.96
Average	42.14/0.97	42.85/0.97	43.11/097	43.88/0.97

Table 2. Comparison of the avrage run-time.

	Wang[11]	Ye[10]	Liu[8]	Our
Run-Time	5.74s	4.58s	2.45s	1.27s

ACKNOWLEDGMENT

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education under Grant 2020R1I1A3A04037680.

References

[1] Wang Y, Wu T, Yang J, Wang L, An W, Guo Y, editors. DeOccNet: Learning to see through foreground occlusions in light fields. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision; 2020.

[2] Zhang J, Liu Y, Zhang S, Poppe R, Wang M. Light field saliency detection with deep convolutional networks. IEEE Transactions on Image Processing. 2020;29:4421-34.

[3] Georgiev TG, Lumsdaine A. Focused plenoptic camera and rendering. Journal of electronic imaging. 2010;19(2):021106.

[4] Wu G, Zhao M, Wang L, Dai Q, Chai T, Liu Y, editors. Light field reconstruction using deep convolutional network on EPI. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017.

[5] Hu Z, Chung YY, Ouyang W, Chen X, Chen Z. Light field reconstruction using hierarchical features fusion. Expert Systems with Applications. 2020;151:113394.

[6] Salem A, Ibrahem H, Kang H-S. RCA-LF: Dense Light Field Reconstruction Using Residual Channel Attention Networks. Sensors. 2022;22(14):5254.

[7] Wang Y, Liu F, Zhang K, Wang Z, Sun Z, Tan T. High-fidelity view synthesis for light field imaging with extended pseudo 4DCNN. IEEE Transactions on Computational Imaging. 2020;6:830-42.

[8] Liu D, Huang Y, Wu Q, Ma R, An P. Multi-angular epipolar geometry based light field angular reconstruction network. IEEE Transactions on Computational Imaging. 2020;6:1507-22.

[9] G. Wu, Y. Liu, L. Fang, and T. Chai, "Revisiting light field rendering with deep anti-aliasing neural network," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

[10] Yeung, HWF, et al. "Fast light field reconstruction with deep coarse-to-fine modeling of spatial-angular clues." Proceedings of the European Conference on Computer Vision (ECCV). 2018.

[11] Wang, F. Liu, Z. Wang, G. Hou, Z. Sun, and T. Tan, "End-to-end view synthesis for light field imaging with pseudo 4DCNN," in Proc. Eur. Conf. Comput. Vis. New York, NY, USA: Springer, Sep. 2018, pp. 333-348.

[12] N. K. Kalantari, T.-C. Wang, and R. Ramamoorthi, ''Learning-based view synthesis for light _eld cameras,'' ACM Trans. Graph., vol. 35, no. 6, p. 193, 2016.