

Light Field Angular Super-Resolution using Convolutional Neural Network with Residual Network

Dong-Myung Kim, Hyun-Soo Kang, Jang-Eui Hong, Jae-Won Suh

College of Electrical & Computer Engineering

Chung-Buk National University

Chungju, Korea

dmkim@cbnu.ac.kr, hskang@cbnu.ac.kr, jehong@cbnu.ac.kr, sjwon@cbnu.ac.kr

Abstract— Light-field images are used in many areas because of its advantage that being able to film multiple images at once, but they are limited in their actual use due to their low resolution. To overcome this drawback, Light-field super-resolution studies are actively being conducted to expand the resolution of light-field images. Light-field super-resolution are divided into two categories. The first is to increase the spatial resolution and the other one is to increase angular resolution. In this paper, we propose an angular super-resolution algorithm based convolution neural network (CNN). Experimental results show that the proposed algorithm produces higher PSNR and SSIM values than existing methods.

Keywords— *light-field; super-resolution; deep learning*

I. INTRODUCTION

Light-field image provides independent brightness information about the direction of light as well as cumulative brightness at each image view point. Therefore, light-field image itself contains 3D geometric information of the scene and uses this additional information of light to enable a variety of applications, such as optical rendering of the virtual view point, virtual refocusing of the camera, stereoscopic display, object insertion and removal, etc. However, the spatial resolution and angular resolution of the light-field camera is limited due to the micro-lens array.

Due to the spatial limitation of the micro-lens array, spatial resolution has to be reduced in order to increase angular resolution. It means that angular resolution and spatial resolution are in a trade-off relationship. For the effective application of the light-field camera, it is necessary to increase the resolution for spatial and angular. This problem can be solved by using super-resolution algorithms that converts low resolution images to high resolution images using rich structural information inherent in light-field images.

Light-field super-resolution algorithms are divided into two categories: spatial and angular super-resolution. The recently studied spatial super-resolution algorithms are CNN-based single image super resolution (SISR) methods [2], [3], [4], [6], [7]. SRCNN [2] is the first end-to-end CNN-based SISR method. The network directly learns an end-to-end mapping

between low- and high-resolution images. SRCNN consists of three convolution layers that represent patch extraction, non-linear mapping, and reconstruction. This process is equivalent to traditional sparse-coding-based super-resolution. VDSR [3] significantly increased the learning rate to train the network with much more layers than SRCNN. Therefore, the network has vanishing and exploding gradients problem due to the many weight layers. To solve the problem, they adopt residual learning and gradient clipping. EDSR [4] employs ResNet [5] architecture called as residual block (ResBlock) to solve degradation problems, such as gradient vanishing or exploding by performing the feedforward method. EDSR network has ResBlock removed batch normalization reduces the flexibility of the network because it normalizes the weights and last rectified linear unit (ReLU). DBPN [6] introduces the up- and down-projection unit that is interconnected differently from recently proposed deep super-resolution networks. Densely connected up- and down-projection unit helps to extract feature maps. The recent trend of SISR research is focused on restoring perceptual quality rather than the absolute quality such as peak signal-to-noise ratio (PSNR). SRGAN [7] use the concept of generative adversarial networks (GANs) [8]. Network minimize perceptual loss that consist of adversarial loss and VGG [9] based content loss. So far, the proposed algorithms have shown good performance both quantitatively and qualitatively.

Angular super-resolution algorithms are also based on CNN. Kalantari [10] proposed disparity estimator and color predictor. Disparity estimator estimates disparity at the novel view and then use this disparity to warp all the input views to the novel view. Color predictor uses all the warped images to generate the final image. LFCNN [11] generates spatial and angular SR images through SRCNN-based networks. Their angular network consists of a total of three layers, with the first parallel layer extracting feature map of novel view and the other layers sharing the weight create novel view. However, this algorithm has disadvantages in that it cannot properly extract features of novel view because it predicts initial features using only three layer. To solve this problem, we add convolution layers to extract the feature map of the novel view, and adopt modified ResBlock to avoid gradient vanishing or exploding.

This research was supported by Basic Science Research Program through the NRF funded by the Ministry of Education(2017R1D1A3B03034476) and IITP grant funded by the Korean government(MSIT) (2016-0-00009).

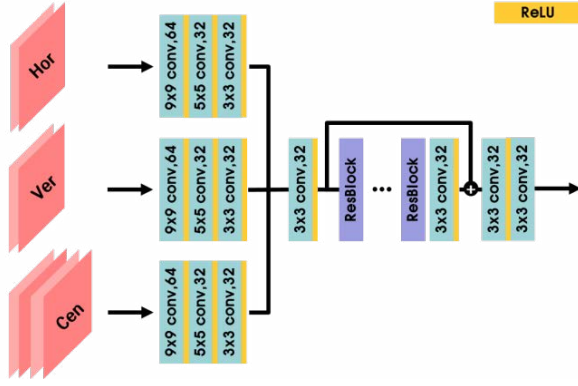


Fig. 1. Architecture of the proposed network

In this paper, we propose end-to-end network that produce novel views of light-field image for enhancing angular resolution. In Section 2, we explain the proposed algorithm, in Section 3, we demonstrate the superiority of the proposed algorithm and conclusion is presented in Section 4.

II. PROPOSED ALGORITHM

The proposed network aims to create a new view point image by receiving sub-aperture images in the light-field image.

As shown in Fig.1, the network first receives each horizontal, vertical, and center pair image in parallel. A pair of input images passes through their convolution layers to generate a feature map, and the generated feature map is input to network sharing the weight. Shared network passes feature maps through the modified ResBlocks and finally generates a new view point of image.

A. Network Architecture

The network of the proposed light-field angular super-resolution is composed as follows. The horizontal, vertical, and central pairs of image are concatenated for using as input. Our network framework is similar to LFCNN [11] but our network is to extract the feature map increases the number of convolution layer. Each parallel network consists of three convolution layers and extracts the feature map of novel view point of image between different pair of images. The first convolution layer creates 64 feature maps extracted from 9×9 filter. The second convolution layer creates 32 feature maps extracted from 5×5 filter. The third convolution layer creates 32 feature maps extracted from 3×3 filter. There is a ReLU behind each convolution layer, and stride is 1. The reason why the filter size of the first layer is larger than the other layers is to consider the disparity information inherent in multiple images.

Extracted feature maps pass through the convolution layers sharing weight. This is for the efficiency of the network, which reduces the number of parameters and increases the performance of the network. The feature maps generated through the previous three convolution layers passes through the convolution layer that creates 32 feature maps extracted from 3×3 filter.



Fig. 2. Modified ResBlock

Generally, as the number of layers in the network increases, the performance of the network increases. However, as the network deepens, problems occur like as gradient vanishing or exploding. We solve the problem by adopting ResBlock. In the network, there are 9 ResBlocks inspired by EDSR [4]. As shown in Fig.2, the difference between the ResBlock structure of EDSR is that it have two additional convolution layer that creates 32 feature maps extracted from 5×5 filter. This improves the performance of the network rather than extracting features with only 3×3 filters. Place the ReLU activation function after the second convolution layer and add the scale module in last convolution layer to stabilize the training procedure. As increased the number of layers in ResBlock, there is a need to adjust the scale factor. The scale factor is decided as 0.2, which is the optimal scale factor for the experimental results. Below Table I. shows the experimental results in different scale factor.

TABLE I. COMPARISON OF DIFFERENT SCALE FACTOR

Scale Factor	PSNR(dB)
0.1	39.2465
0.2	39.2527
0.3	39.2060

B. Loss Function

Most super-resolution networks use mean square error(MSE) as loss function. However, some SR networks such as EDSR [4] and RDN [12] use mean absolute error (MAE) as loss functions instead of MSE to improve the network's performance. In this paper, we decide loss function as MAE, which is to produce better results than MSE. We train our network by minimizing the loss between ground truth image X and created novel view point image Y. The loss function is defined by

$$Loss = \frac{1}{N} \sum_{i=1}^N |Y_i - X_i| \quad (1)$$

C. Model Training

The proposed network was implemented using Tensorflow. After deciding which batch image to train randomly from inside the dataset, the random area was cropped to 32×32. The RGB image was converted to YCbCr and the luminance channel was used only. Network was trained Adam Optimizer was used for optimization, $\beta_1=0.9$, $\beta_2=0.999$. And the learning rate is initialized as 10^{-4} and decreased to 10^{-6} by 10^{-1} every 5000 epoch.

III. EXPERIMENTAL RESULTS

The Lytro Illum camera dataset [13] was used as the training dataset. Each image consists of a set of 14×14 Light Field images, but only central clean 8×8 images were used for training. Each image has a size of 376×541 . Compared with LFCNN [11], the training time increased 2.12 times per epoch due to the increase in the number of layers. The simulated environment is running on a PC with Intel Core i7-8700 CPU 3.2GHz and Nvidia 2080Ti GPU.

Table II shows test result comparison of PSNR and SSIM. According to these results, the PSNR of proposed algorithm is about 1.83 dB higher than LFCNN [11]. In Fig.3, we provide visual comparison of central novel view result images. LFCNN [11] shows that the details of novel view image are dispersed without being merged into one, and the proposed algorithm reconstructs the details more precisely than LFCNN [11].

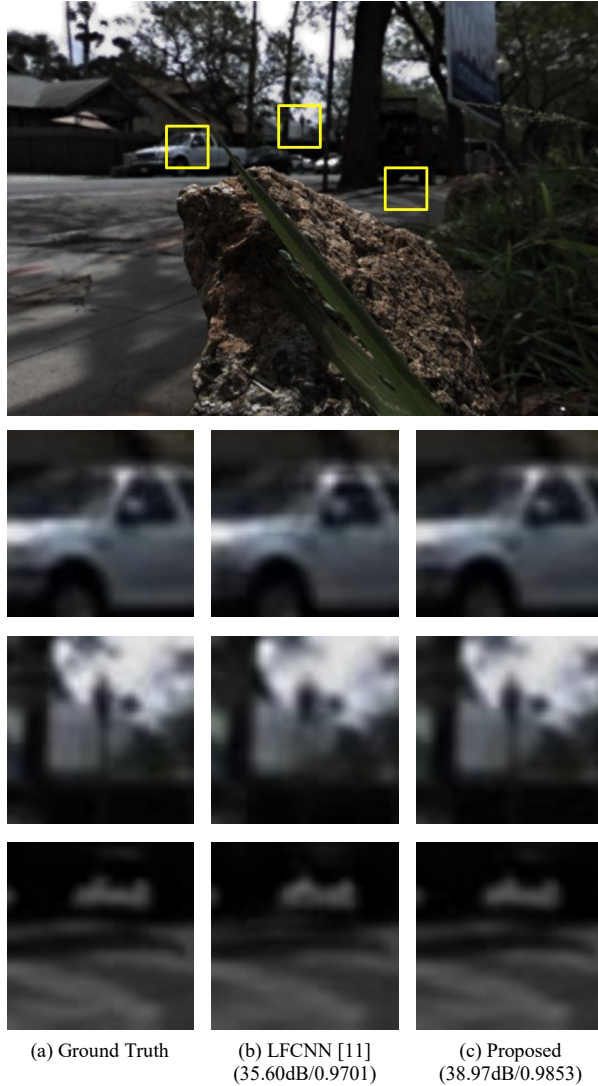


Fig. 3. Visual comparison of central novel view result images.

TABLE II. EXPERIMENTAL RESULTS

Picture Name	LFCNN [11]		Proposed	
	PSNR(dB)	SSIM	PSNR(dB)	SSIM
Cars	36.2069	0.9877	37.0935	0.9897
Flower 1	38.3258	0.9896	39.7853	0.9928
Flower 2	37.8460	0.9847	39.1634	0.9894
Rock	36.0847	0.9721	40.1232	0.9872
Seahorse	38.8997	0.9891	40.0983	0.9913
Average	37.4726	0.9847	39.2527	0.9901

IV. CONCLUSION

In this paper, a new light-field angular super-resolution method was proposed using the CNN with modified residual block. The proposed algorithm was based on the LFCNN [11] network framework. We added convolution layers to solve the problem of LFCNN, which has a small number of layers and is insufficient to feature extraction. As the network deepens, occurred gradient vanishing or exploding problems, which can be solved by adopting modified ResBlocks. The results of the experiment show that the results were better than the existing methods.

REFERENCES

- [1] S. Moitra, "Single-Image Super-Resolution Techniques: A Review," IJSART, Vol. 3, Issue 4, April 2017.
- [2] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," In ECCV 2014.
- [3] J. Kim, J. Kwon Lee, and K. M. Lee, "Accurate image superresolution using very deep convolutional networks", In CVPR 2016.
- [4] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," In CVPR 2017.
- [5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," In CVPR 2016.
- [6] M. Haris, G. Shakhnarovich, N. Ukita, "Deep Back-Projection Networks For Super-Resolution," In CVPR 2018.
- [7] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," In CVPR 2016.
- [8] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," In NIPS 2014.
- [9] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", In ICLR, 2015.
- [10] N. K. Kalantari, T.-C. Wang, and R. Ramamoorthi, "Learning-based view synthesis for light field cameras," ACM Trans. Graph., vol. 35, no. 6, p. 193, 2016.
- [11] Y. Yoon, H. Jeon, D. Yoo, J. Lee, I. Kweon, "Light-Field Image Super-Resolution Using Convolutional Neural Network," IEEE Signal Processing Letter, 2017.
- [12] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, "Residual Dense Network for Image Super-Resolution", In CVPR 2018.
- [13] A. S. Raj, M. Lowney, and R. Shah. Light-Field Database Creation and Depth Estimation. Accessed: Jun. 20, 2017. [Online]. Available:<http://lightfield.stanford.edu/lfs.html>
- [14] N Bin Liu, Zhisheng Yan, and Chang Wen Chen, "Medium Access Control for Wireless Body Area Networks with QoS Provisioning and Energy Efficient Design," IEEE Transactions on Mobile Computing, Vol. 16, No. 2, Feb. 2017, pp. 422-434.