Light Field Synthesis from a Single Image using Improved Wasserstein Generative Adversarial Network

Lingyan RUAN [1], Bin CHEN[1] and Miu-Ling LAM[1,2]

[1]City University of Hong Kong, Hong Kong
[2] CityU Shenzhen Research Institute

Abstract

We present a deep learning-based method to synthesize a 4D light field from a single 2D RGB image. We consider the light field synthesis problem equivalent to image super-resolution, and solve it by using the improved Wasserstein Generative Adversarial Network with gradient penalty (WGAN-GP). Experimental results demonstrate that our algorithm can predict complex occlusions and relative depths in challenging scenes. The light fields synthesized by our method has much higher signal-to-noise ratio and structural similarity than the state-of-the-art approach.

CCS Concepts

•***Computing methodologies*** → ***Machine learning;*** *Computational photography;*

1. Introduction

Light field synthesis refers to the generation of a dense collection of novel views from a single view. It offers great opportunities for applying the powerful features of light field photography, including depth refocusing and aperture adjustment, to conventional 2D photographs.

In this paper, we present a learning-based view synthesis algorithm to generate a 4D light field from a single RGB image. The key idea is to treat the light field synthesis problem as image super-resolution, and then utilize a GAN-based framework to upscale a low-resolution input (RGB image) to a high-resolution light field image. The goal of this work is similar to [SWS*17], which proposed a learning-based method consisting of two convolutional neural networks (CNNs), one estimates scene depth and renders Lambertian light field, while the other predicts the occluded rays and non-Lambertian effects. However, the approach suffers from the defective depth map estimation which leads to severe artifacts and failure of view synthesis. On the contrary, our method does not require explicit estimation of depth map, thus avoiding failure caused by inaccurate depth information, thus is capable of generating more convincing and robust light fields.

2. Method

The network architecture used in this paper is similar to the GAN for image super-resolution (SRGAN) [LTH*16], but WGAN-GP [GAA*17] is used instead of DCGAN, and 8× upscaling factor is used instead of 4×. We used the light field image dataset provided by [SWS*17], which includes 3275 flowers images captured with the Lytro Illum camera. Each light field has 14×14 angular views. However, many angular samples at the corners are not within the camera's aperture, so only 8×8 grid of angular samples lie within the aperture were used in this work. We feed the central 2D slice of a light field as the input and upscale it 8× to yield a high-resolution image. The generator network is trained to mimic the manifold of the real light field image and generate the same size high-resolution image. Meanwhile, the discriminator network is trained to distinguish the generated high-resolution images from the real light filed images. We formulate the mixed loss function as:

$$L = L_{wgan-gp} + \lambda_{mse} L_{mse} + \lambda_{vgg} L_{vgg} + \lambda_{vgg\_epi} L_{vgg\_epi} \quad (1)$$

We introduce both pixel-wise loss (MSE) and perceptual loss similar to SRGAN in [LTH*16], where the perceptual loss is defined based on the pre-trained 19 layer VGG network [SZ14] within feature representations between the generator image and reference image. Moreover, we add the perceptual loss for image on epipolar plane to ensure the depth information can be generated properly.

3. Results

We randomly divide the dataset of 3275 light fields into three groups: 2800 for training, 395 for cross-validation and 80 for testing. The network was trained with a NVIDIA GeForce GTX 1080 graphic card. We benchmark the performance of our method with [SWS*17]. The results of qualitative and quantitive performance comparison are shown in Figure 1 and Figure 2 respectively.

Figure 1 shows the qualitative comparison between [SWS*17]

DOI: 10.2312/egp.20181017

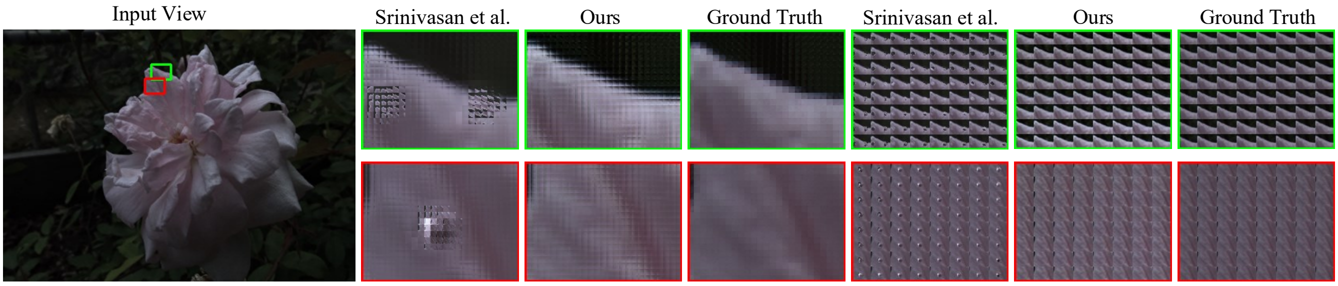

Figure 1: *We compared our results against [SWS*17]. Three columns in the middle represent the same part of synthetic light fields using the methods by [SWS*17], by us and the ground truth. Three columns on the right are the corresponding extracted 64 novel views* (8×8).

and our method. Notable artifacts can be observed in the light fields generated by [SWS*17], while our results look closer to the ground truth.

Figure 2 shows the quantitive comparison between [SWS*17] and our method. The peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) of all 80 testing images are computed for evaluation. The histograms demonstrate that our approach outperforms in terms of both PSNR and SSIM.

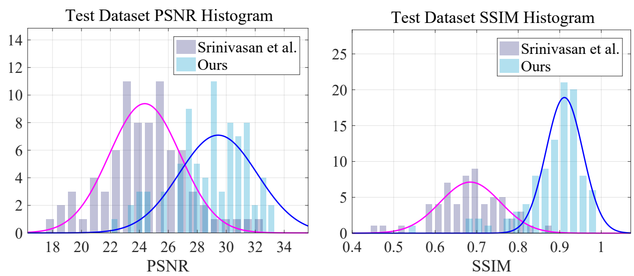

Figure 2: *Histograms of peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) of 80 testing images for quantitative evaluation.*

Figure 3 shows the light fields synthesized by our network as corner view crops. The epipolar slice crops illustrate that our algorithm can predict complex occlusions and relative depths. The figure also demonstrates that our synthesized light fields can perform convincing refocusing.

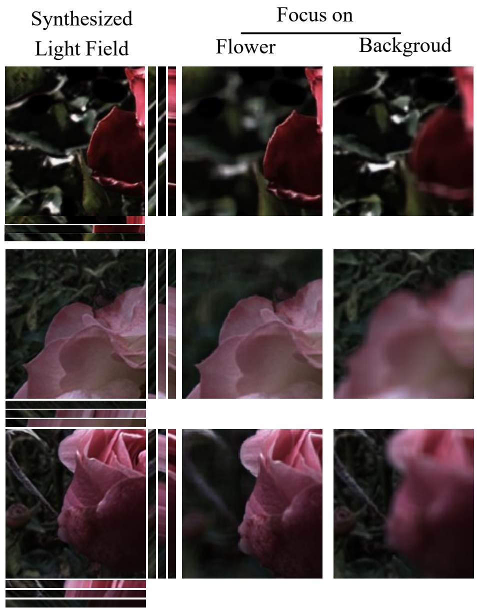

Figure 3: *We present our synthesized light field as a corner crop with the corresponding epipolar slice image. The lines at different slopes shows that our GAN-based algorithm is able to learn complicated occlusion. Moreover, we successfully refocus the image from foreground to background.*

4. Conclusions

In this paper, we introduce a novel method to generate light field from a single 2D image directly by using the improved WGAN, which can be easily applied to everyday photography. Our method provides a simple but powerful approach to synthesize light fields without depth map estimation.

5. Acknowledgements

This work was partially supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (CityU 11206216), National Natural Science Foundation of China (61502406), Croucher Foundation (9500016) and ACIM-SCM.

References

[GAA*17] GULRAJANI I., AHMED F., ARJOVSKY M., DUMOULIN V., COURVILLE A.: Improved training of wasserstein gans. *arXiv preprint arXiv:1704.00028* (2017). 1

[LTH*16] LEDIG C., THEIS L., HUSZÁR F., CABALLERO J., CUNNINGHAM A., ACOSTA A., AITKEN A., TEJANI A., TOTZ J., WANG Z., ET AL.: Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint arXiv:1609.04802* (2016). 1

[SWS*17] SRINIVASAN P. P., WANG T., SREELAL A., RAMAMOORTHI R., NG R.: Learning to synthesize a 4d rgbd light field from a single image. In *International Conference on Computer Vision (ICCV)* (2017), vol. 31, p. 6. 1, 2

[SZ14] SIMONYAN K., ZISSERMAN A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014). 1