Learning a Deep Convolutional Network for Light-Field Image Super-Resolution

Youngjin Yoon yjyoon@rcv.kaist.ac.kr
Hae-Gon Jeon hgjeon@rcv.kaist.ac.kr
Donggeun Yoo dgyoo@rcv.kaist.ac.kr
Joon-Young Lee jylee@rcv.kaist.ac.kr
In So Kweon iskweon@kaist.ac.kr

Robotics and Computer Vision Lab., KAIST

Abstract

Commercial Light-Field cameras provide spatial and angular information, but its limited resolution becomes an important problem in practical use. In this paper, we present a novel method for Light-Field image super-resolution (SR) via a deep convolutional neural network. Rather than the conventional optimization framework, we adopt a data-driven learning method to simultaneously up-sample the angular resolution as well as the spatial resolution of a Light-Field image. We first augment the spatial resolution of each sub-aperture image to enhance details by a spatial SR network. Then, novel views between the sub-aperture images are generated by an angular super-resolution network. These networks are trained independently but finally fine-tuned via end-to-end training. The proposed method shows the state-of-the-art performance on HCI synthetic dataset, and is further evaluated by challenging real-world applications including refocusing and depth map estimation.

1. Introduction

Recently, light-Field (LF) imaging introduced by Adelson and Bergen [1] has come into the spotlight as the next generation camera. A LF camera is able to acquire spatial and angular information of light ray distribution in space, therefore it can capture a scene from multiple views in a single photographic exposure.

Starting from Lytro [19] and Raytrix [21], commercial LF cameras have demonstrated their capabilities to exhibit refocusing and 3D parallax from a single shot. With richer angular information in a LF image, many studies have shown the potential to improve the performance of many computer vision applications such as alpha matting [6], saliency detection [17], LF panorama [3], and depth reconstruction [27, 33, 13].

Along with such advantages of the LF imaging, several researches also have pointed out that the low spatial and angular resolution of LF images becomes the main difficulty in exploiting its advantage. The LF imaging has a trade-off between a spatial and an angular resolution in a restricted sensor resolution. A micro-lens array, placed between a sensor and a main lens, is used to encode angular information of light rays. Therefore, enhancing LF images resolution is crucial to take full advantage of LF imaging.

For image super-resolution (SR), most conventional way is to perform optimizations with prior information [16, 26]. The optimization-based approach shows many promising results, however it generally requires parameter tuning to adjust the weight between a data fidelity term and a prior term (e.g., local patch sizes, color consistency parameters, etc.). Away from the optimization paradigm, recently data-driven learning methods based on deep neural network models has been successfully applied to image SR [8, 11]. One major benefit of these learning approaches is their generalization ability. If training data is sufficiently large to fit a model and the data covers a wide range of distributions of expected test images, we can expect generalized performance even without careful engineering to control the distributions.

The goal of this paper is to obtain a set of high-resolution LF images with a data-driven supervised learning approach. Here, the SR target includes the number of sub-aperture images as well as the number of spatial pixels. Similar to the CNN-based SR method, proposed by Dong *et al.* [11], we also adopt a deep convolutional neural network (CNN) [9] and solve this problem by CNN-based regression with an Euclidean cost.

There are different lines of work that adopt deep learning frameworks to solve image restoration problems, including SR [11, 8], denoising [12] and deblurring [32, 25]. Compared to them, our major focus lies on LF imaging, which is the specific domain having different restoration problems and applications. While the previous learning approaches deal with the images captured from standard cameras, this paper is, to our knowledge, the first successful trial that applies the CNN framework to the domain of LF images. Our method, named Light-Field Convolutional Neural Network (LFCNN), augments the number of views as well as spatial resolutions for further benefits in accurate depth estimation and refocusing.

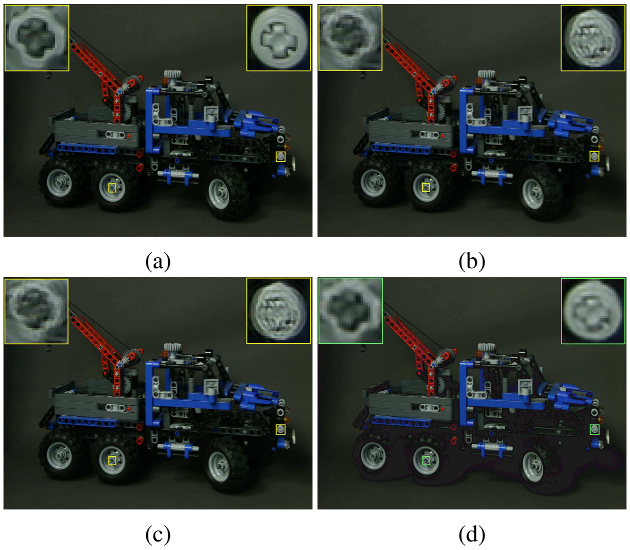

(a) (b) (c) (d)

Figure 1. (a) Ground-truth sub-aperture image. (b) An averaged image of adjacent sub-aperture images. (c) Each of sub-aperture images in (b) is upsampled by SRCNN [11], then averaged. (d) Our result.

2. Related Work

Deep learning for image restoration Recently, the deep learning approach has been successfully applied to the image restoration problems such as SR [8, 11], denoising [2, 12, 22] and image deblurring [25, 32]. Schuler *et al*. [22] show that a multi-layer perceptron can solve an image deconvolution problem. Agostinelli *et al*. [2] combine multiple denoising autoencoders to handle various types of noise. Several CNN models also have been addressed to solve image restoration. Eigen *et al*. [12] propose a CNN to remove raindrops and lens dirts. Xu *et al*. [32] propose a cascaded CNN for non-blind deblurring. The first network is for deconvolution and the second network is built based on [12] to remove outliers. Sun *et al*. [25] design a CNN for non-uniform motion blur estimation with Markov Random Field.

Our framework is motivated by Dong *et al*. [11] who relate the sparse representation to the stacked nonlinear operations to design a CNN for a single image SR. Compared to [11], our LFCNN jointly solves the spatial and angular SR specialized for LF imaging. Therefore, our method presents much accurate high-resolution novel views than the traditional multi-dimension interpolations and optimization-based methods as shown in Fig. 1.

SR for Light-Field imaging Since hand-held LF cameras are released, LF images suffer from a lack of spatial and angular resolution due to a limited sensor resolution. Bishop and Favaro [4] present a Bayesian inference technique based on a depth map for a LFSR. Cho *et al*. [7] present a method for Lytro image decoding and sub-aperture images generation. They also propose a dictionary learning based sub-aperture image SR. Shi *et al*. [24] present LF signals reconstruction using sparsity in continuous frequency domain. The sparse signal can be recovered by optimizing frequency coefficients using a 2D sparse Fourier transform reconstruction algorithm. These approaches only up-sample either spatial or angular resolution of a LF image, which limits the performance of LF applications such as a realistic digital refocusing and depth estimation.

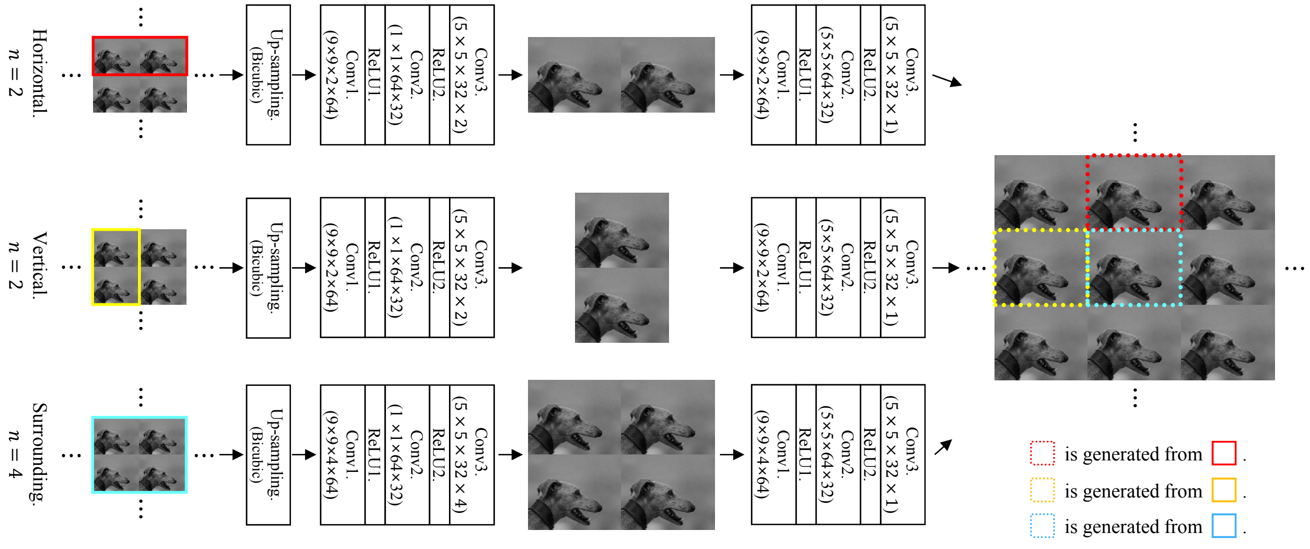
There are recent studies to achieve spatial and angular SR simultaneously. Wanner and Goldluecke [29] introduce a variational LFSR framework by utilizing the estimated depth map from Epipolar plane image. Mitra and Veeraraghavan [20] propose a patch-based approach using a Gaussian Mixture Model prior and a inference model according to disparity of a scene. However, the low quality LF images captured by commercial LF cameras degrade the performance of these approaches.

We refer readers to [18] for a detailed description about theoretical analysis on the source of the achievable resolution given by micro-lens LF cameras. In [18], the authors mention that a prefilter kernel for LF rendering should vary as depth-dependent and spatially-variant. The uniqueness of our network is that specific networks for angular and spatial resolution enhancement have the merit of sub-aperture images recovery regardless of scene depth and spatial variance. As will be demonstrated in our experiments, our framework is highly effective and has outperformed the state-of-the-art algorithms in LFSR from a lenslet LF image.

3. Limitations of EPI Processing

One major benefit of LF images over conventional images is access to Epipolar plane image (EPI) which is 2D slices of constant angular (vertical resolution) and spatial direction (horizontal resolution). As the EPI is only consisted of lines with various slopes, it makes image processing and optimization tractable and previous work [30] performed LFSR on the EPI. However, we have found two practical problems that we cannot use EPIs for LFSR using CNN frameworks, especially when we consider commercial LF cameras.

While the resolution to the spatial axis is sufficient, the number of sub-aperture images are too few (e.g. 9×9 or 5×5) to suitably apply multiple convolutional layers for sub-aperture SR. This becomes a serious problem because we aim to enhance both angular and spatial resolution of a LF simultaneously. Performing convolutions across such a small axis results in loss of substantial portion of angular information in both ends edge, whereas this loss in the spatial axis is minor due to their enough resolution.



Spatial SR networks. Angular SR networks.

Figure 2. Overview of our LFSR framework. Stacked input images are up-scaled to a target resolution by the bicubic interpolation, and are fed to the spatial SR network. Each color box represents a horizontal, a vertical and a surrounding input pair, respectively. The output of the spatial SR network is used as the input of the angular SR network. The angular SR network produces high-resolution novel views.

The other problem is that LF images captured by commercial LF cameras suffer from severe artifacts such as lens aberration, microlens distortion and vignetting as discussed in [14]. The artifacts have a negative impact on EPIs. When we capture a planar scene, we observe spatially-variant EPI slopes due to the artifacts. The degree of the distortion additionally varies for each sub-aperture image. A micro-lens array, placed between sensor and main lens, to encode angular information of rays also raise the problem. The microlens array occludes lights and leads to a reduced signal-to-noise ratio. Even with only a few pixels near EPI slopes which are corrupted by the noise, processing on EPIs may fail to show promising results. Therefore, a practical way to super-resolve LF images should be considered to solve the issue which we should address now in order to take full advantage of LF images.

4. Joint SR for Light-Field imaging

4.1. Overview

We propose a new deep learning structure, which jointly increases the resolution in both the spatial domain and the angular domain. Fig. 2 is an overview of our LFCNN model composed of a spatial SR network and an angular SR network. The spatial SR network, which is similar to [11], enhances spatial resolution, and the angular SR network augments the number of sub-aperture images. A pair of neighboring sub-aperture images is initially interpolated by a desired up-scaling factor, then is fed to the spatial SR network to recover high-frequency details. The output of the spatial SR network is fed to the angular SR network which produce a novel view between the input sub-aperture images. The reason why we first place the spatial SR network followed by the angular SR network will be discussed in Sec. 4.4. In the example shown in Fig. 2, we obtain three up-sampled images where one of them is a novel view generated from the pair of input sub-aperture images.

4.2. Spatial SR network

Our spatial SR network is similar to [11] but takes and outputs dual-images ($n = 2$) or quad-images ($n = 4$), because multiple images are required to generate a novel view between them. As illustrated in Fig. 2, the spatial SR network is composed of three convolution layers, and the layers are composed of 64 filters of $9 \times 9 \times n$ size, 32 filters of $1 \times 1 \times 64$ size, and n filters of $5 \times 5 \times 32$ respectively. The first two layers are followed by ReLU layer. We feed input images to the network after we initially interpolate the images by a desired upsampling factor (×2 in this paper).

4.3. Angular SR network

A simple way to produce a novel view between two neighboring sub-aperture images is to apply the bilinear interpolation to the sub-aperture images. Such simple interpolation introduces the loss in high-frequency information and results in blurred edges. Wanner and Goldluecke [29] additionally use an accurate depth map as a geometric guidance and perform angular SR. However, in practice, LF images from commercial LF cameras have very narrow baseline, therefore it is very difficult to estimate a reliable depth map and to utilize the estimated depth map as a guidance.

To tackle these difficulties, we employ a data-driven supervised learning approach because numerous training samples can be easily formed. For example, let us consider a row of a LF image array. The i-th sub-aperture image can be regarded as a ground truth of a novel view between the $(i-1)$-th and $(i+1)$-th images. We then can train a mapping function from sub-aperture images to a novel view between them via stacked non-linear operations. As the mapping function, we adopt a CNN composed of three convolution layers, named angular SR network, as depicted in Fig. 2. This network takes n images and produces one novel view. The first, second and third convolution layers are composed of 64 filters of $9 \times 9 \times n$ size, 32 filters of $5 \times 5 \times 64$ size, and a filter of $5 \times 5 \times 32$, respectively. The first and second layers are followed by ReLU layer.

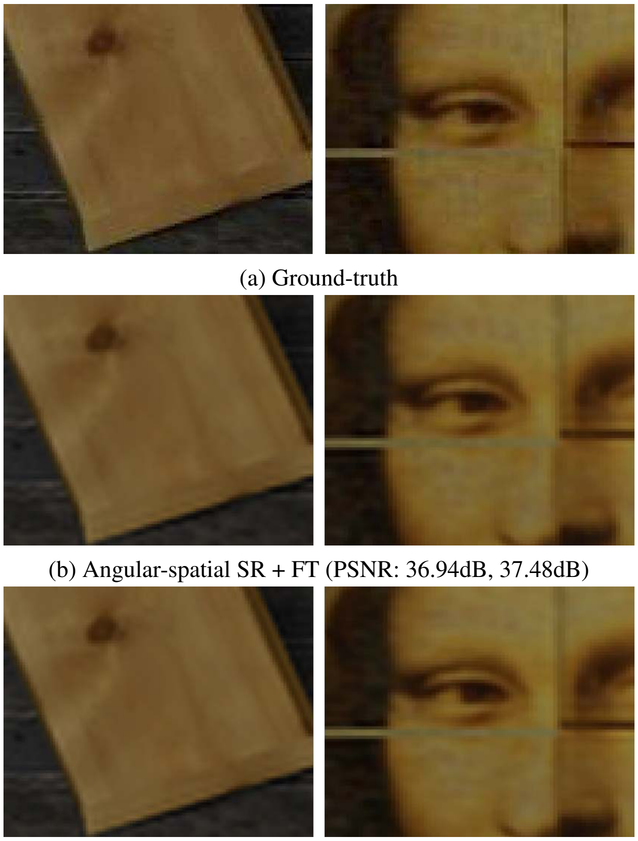
To fully augment a given $(M \times M)$-view array into a $(2M-1) \times (2M-1)$-view array, we compose three angular SR networks taking three different types of input: a horizontal pair ($n = 2$), a vertical pair ($n = 2$) and surroundings ($n = 4$). The surrounding type takes four sub-aperture images located in each corner of a target view. The three networks has the same architecture except the depth n of filters in the first convolution layer.

4.4. Spatial-angular SR vs angular-spatial SR

In our cascaded SR strategy, we are able to consider the angular-spatial SR scheme as well as the spatial-angular scheme of Fig. 2. These two options also show quite similar performances according to the evaluation (Table 1) on the synthetic HCI dataset. However, we empirically observe that, in the real-world experiment, the spatial-angular SR shows better performance in general compared to the angular-spatial SR. As shown in Fig. 4, the spatial-angular SR produces much shaper results, while the angular-spatial SR produces blurred images. Since low-resolution sub-aperture images are fed into the angular SR network first in case of the angular-spatial scheme, the details in low resolution images may not be well-localized in a novel view, therefore the details in the novel view result in inaccurate edges and points through the spatial SR network. On the other hand, the spatial-angular SR scheme produces clean images in general because the localization ambiguity is reduced.

5. Training LFCNN

To train the spatial SR network, we need a bunch of pairs of blurred input and sharp ground-truth patches. Following [11], we synthesize blurred images from original images. We first down-sample original images and up-sample them again using bicubic interpolation to produce the blurred version of original images. Then, the patch from an original image is regarded as a ground-truth and the patch from the corresponding blurred image becomes an input. From the pairs of blurred and sharp LF images, we randomly crop $44 \times 44 \times n$ size patches in the same region of n neighboring sub-aperture images. The output patch is reduced to $32 \times 32 \times n$ dimensions because of stacked convolution operations with spatial filters.


(a) Ground-truth

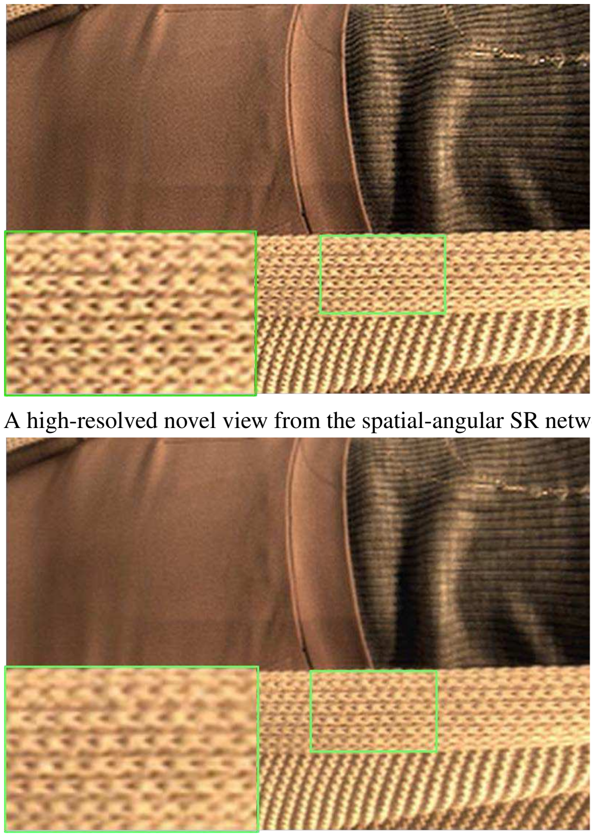
(b) Angular-spatial SR + FT (PSNR: 36.94dB, 37.48dB)

(c) Spatial-angular SR + FT (PSNR: 36.92dB, 37.64dB)

Figure 3. Qualitative comparison according to the order of SR networks on synthetic images.

In the angular SR network, the size of an input patch should be same to the output dimension ($32 \times 32 \times n$) of the previous spatial SR network. We therefore randomly crop $32 \times 32 \times 1$ size patches in the same region from n surrounding sub-aperture images for input patches and from the corresponding central sub-aperture images for ground-truth patches.

We use Caffe[15] to implement and train our LFCNN. The spatial and angular SR networks are initialized by random weights with a Gaussian distribution of a zero mean and a standard deviation of 10^{-3}. With an Euclidean cost, these networks are trained independently but finally fine-tuned in an end-to-end manner. For end-to-end fine-tuning, we simply connect the last layer of the spatial SR network



(a) A high-resolved novel view from the spatial-angular SR network

(b) A high-resolved novel view from the angular-spatial SR network

Figure 4. Qualitative comparison according to the order of SR networks on a real-world SR image.

to the first layer of the angular SR network. In this case, $44 \times 44 \times n$-dimensional input is fed to the whole network, and the network is supervised with $32 \times 32 \times 1$-dimensional ground-truth patches. Following [11], the learning rate is set to 10^{-4} for the first two layers and 10^{-5} for the last layer in each network. For the end-to-end fine-tuning, we decrease the learning rates to 10^{-5} and 10^{-6}.

6. Experiments

We evaluate our LFCNN on both synthetic real-world datasets. We also apply our LFCNN to the challenging real-world applications such as refocusing and depth-map estimation.

6.1. Synthetic dataset

To compare our method with the state-of-the-art LFSR approaches [20, 29], we select "Buddha" and "Mona" in the HCI dataset [31] as test images because these two images have fine structures and large depth variations. The remaining 10 images in the HCI dataset are used for generating patches to train LFCNN. Given this small dataset having only 12 images, it is the best choice for us to secure 10 images as training data since it is preferred to have large numbers of training data for the best performance of our CNN-based approach.

We train and test our LFCNN in the following SR setting: spatial SR from a 384×384 size image to a 768×768 size image, and angular SR from a 5×5 size LF array to a 9×9 size LF array. For training, we extract 1,200,000 patches of 44×44 size from the training images. We train each network by 10^8 iterations and fine-tune the whole network by 25,000 iterations.

Table 1 shows the result of quantitative evaluation on the HCI dataset. For the evaluation, we compare our method with the state-of-the-art methods [20, 29] as well as the standard interpolation methods as baselines (4D bilinear and 4D bicubic). The results of [20, 29] are obtained by the source codes provided by the authors [1], where we carefully tune the parameters to maximize performance. To analyze the performance according to the architectural schemes of our method, we place the angular SR network at first followed by the spatial SR network, and vise versa. For each scheme, we also fine-tune the whole networks in an end-to-end training manner, which is noted by "FT".

The peak signal-to-noise ratio (PSNR) and the gray-scale structural similarity (SSIM) [28] are used as the evaluation metrics. To examine robustness to scene dependency, we measure the PSNR and SSIM scores of all the estimated novel views and report the minimum, average, and maximum values of them for each dataset. As shown in Table 1, our method consistently yields higher PSNR and SSIM scores than the other methods even without fine-tuning. Before end-to-end fine-tuning, the angular-spatial scheme shows the best performance. The end-to-end fine-tuning consistently improves the performance of our approach in general. After fine-tuning, two schemes of our approach show negligible difference and largely outperform the state-of-the-art methods [20, 29]. Fig. 3 shows our results according to the different architectural schemes for qualitative comparison.

6.2. Real-world dataset

Due to the numerous number of parameters in CNN, many previous image restorations methods based on CNN [8, 11, 12, 25, 32] generate training samples from ImageNet after the task-specific image processing. However, in our case, there are insufficient number of public LF images. In addition, the public LF datasets have different angular resolutions, which result in different depth variations across the datasets. Thus, combining these datasets to make one large set of LF images for training is impossible. We therefore took more than 300 scenes having various tex-

[1] The authors of [20] provided their source code to us. For [29], the source code was downloaded at http://hci.iwr.uni-heidelberg.de/HCI/Research/LightField/old_page/

Methods	PNSR(dB)						SSIM					
	Buddha			Mona			Buddha			Mona		
	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
Bilinear	33.57	33.66	33.78	34.14	34.25	34.32	0.9036	0.9151	0.9242	0.9242	0.9291	0.9320
Bicubic	34.22	34.63	35.14	34.10	34.20	34.25	0.9251	0.9334	0.9466	0.9484	0.9496	0.9512
SpatialSR(Bicubic)+AngularSR(ours)	35.68	35.79	35.87	35.80	35.91	35.99	0.9284	0.9293	0.9300	0.9362	0.9364	0.9367
AngularSR(ours)+SpatialSR(ours)	36.54	36.64	36.71	37.10	37.20	37.28	0.9550	0.9557	0.9565	0.9657	0.9659	0.9662
AngularSR(ours)+SpatialSR(ours)+FT	**36.78**	**36.86**	**36.94**	37.31	37.40	37.48	**0.9571**	**0.9580**	**0.9589**	**0.9667**	**0.9669**	**0.9671**
SpatialSR(ours)+AngularSR(ours)	35.76	35.87	35.93	36.25	36.33	36.39	0.9466	0.9475	0.9481	0.9575	0.9578	0.9580
SpatialSR(ours)+AngularSR(ours)+FT	36.71	36.84	36.92	**37.46**	**37.56**	**37.64**	0.9549	0.9558	0.9565	0.9637	0.9640	0.9644
Mitra and Veeraraghavan [20]	22.61	26.76	32.37	24.36	28.11	34.53	0.6105	0.7764	0.9126	0.6328	0.7728	0.9563
Wanner and Goldluecke [29]	21.77	25.50	33.83	25.46	29.62	36.84	0.5251	0.6502	0.9107	0.5977	0.7432	0.9441

Table 1. Quantitative evaluation on the synthetic HCI dataset. Our approach significantly outperforms the state-of-the-art methods.

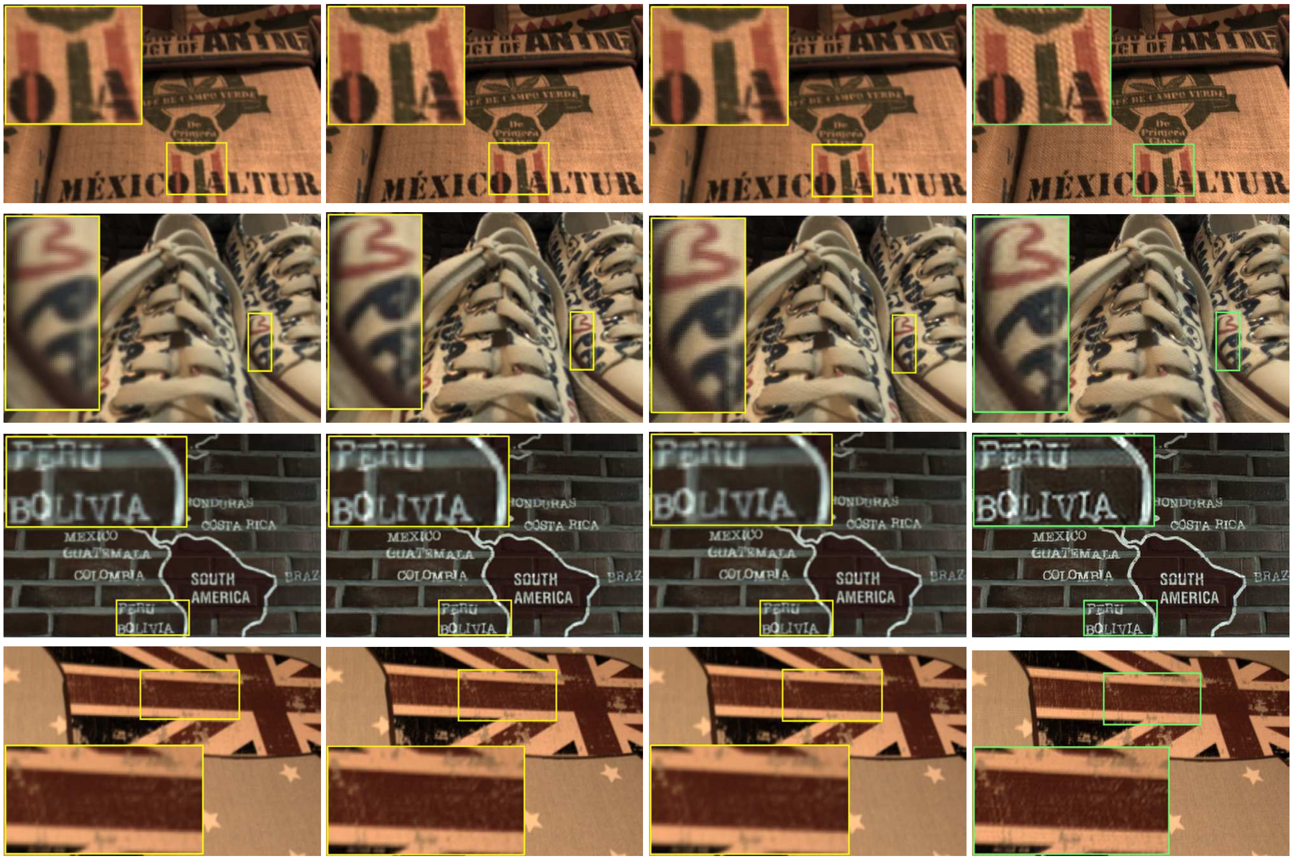
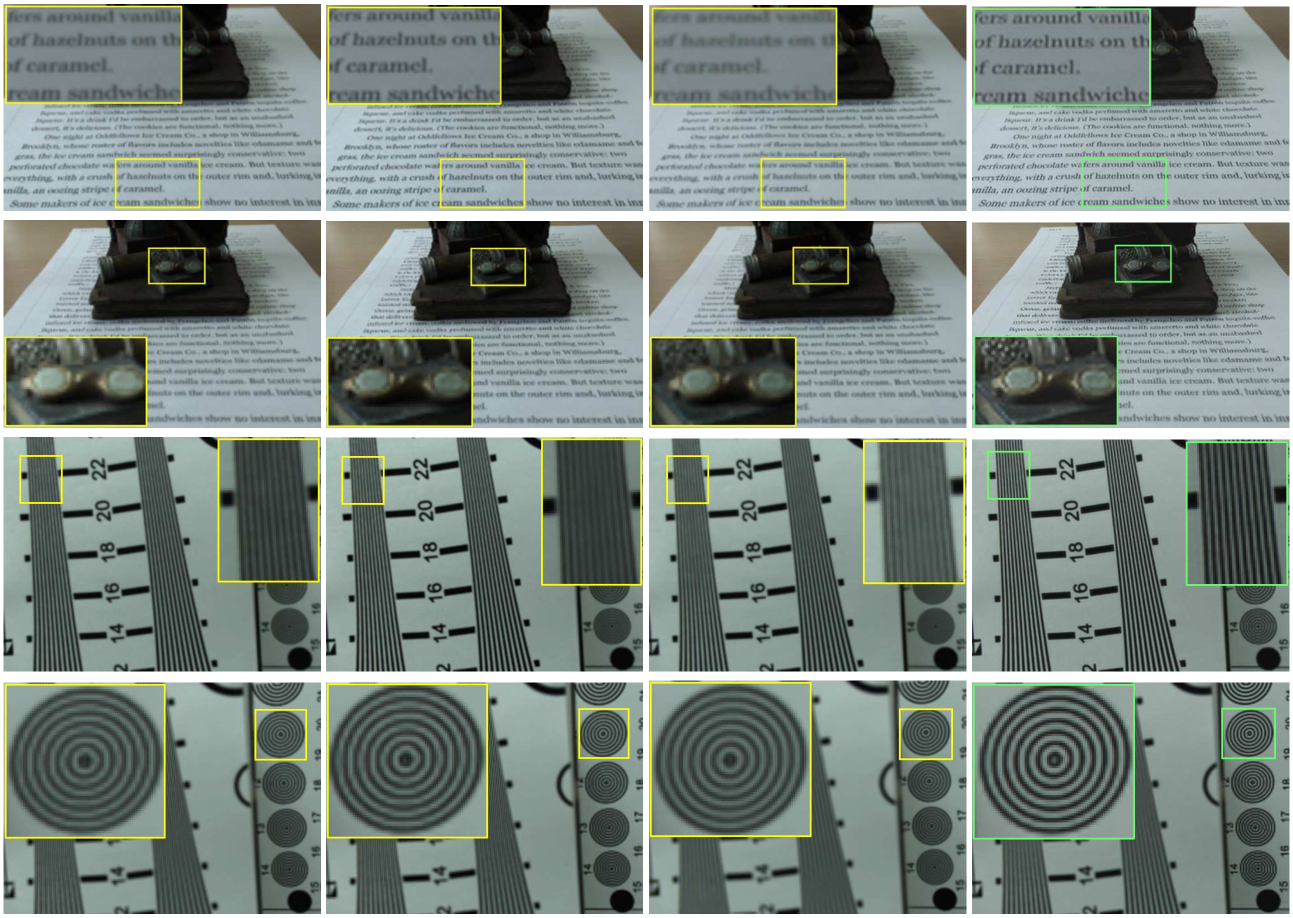

(a) Bilinear (b) Bicubic (c) Mitra and Veeraraghavan [20] (d) Our spatial-angular SR.

Figure 5. Qualitative comparisons on real-world datasets. Input images are enhanced by a factor of 2.

tures and depth with a Lytro Illum camera for training over real-word LF images. We extract 5×5 sub-aperture images with 383×583 spatial resolution from the real LF images using the LF geometric toolbox provided by [5]. We extract 2,000,000 patches of 44×44 and train each network in the same manner as stated in the synthetic experiment.

In Fig. 5, we demonstrate the upsampling results for qualitative evaluation on the real-world dataset because our real-word dataset does not have ground-truth data. The scaling factor is 2 for both of spatial SR and angular SR. For comparison, we also demonstrate the results of the state-of-the-art method [20] and two standard interpolation methods including 4D bilinear and 4D bicubic. For [20], we rigorously tuned the parameters to gain the maximum performance[2]. Both 4D bilinear and 4D bicubic produce blurred

[2] This method is based on EPIs and is sensitive to the disparity pattern which depends on the depth of the scene. There is difficulty in handling LF images captured by LF cameras since the EPIs are noisy and inaccurate.



(a) Bilinear (b) Bicubic (c) Mitra and Veeraraghavan [20] (d) Our spatial-angular SR.

Figure 6. Refocused images after applying different upsampling methods. The images in the first and third row are focused on a nearby object. The images in the second and fourth row are focused on a distant object.

results because these methods fail to interpolate a novel view with a large depth variation. Especially, the nearby regions of the images having larger disparity between adjacent views suffer from inaccurate angular SR. On the other hand, our method produces much sharper results. These results clearly demonstrate that our cascaded learning framework works as well on real-world data.

It is worth to mention that the method proposed by Wanner and Goldluecke [30] completely fail to work on our real-world dataset. The performance of the method [30] highly depends on the quality of estimated disparity maps. As discussed in [27, 14], the EPI-based approach may fail to generate reliable depth maps due to very narrow baseline and severe noise of LF images from a Lytro camera.

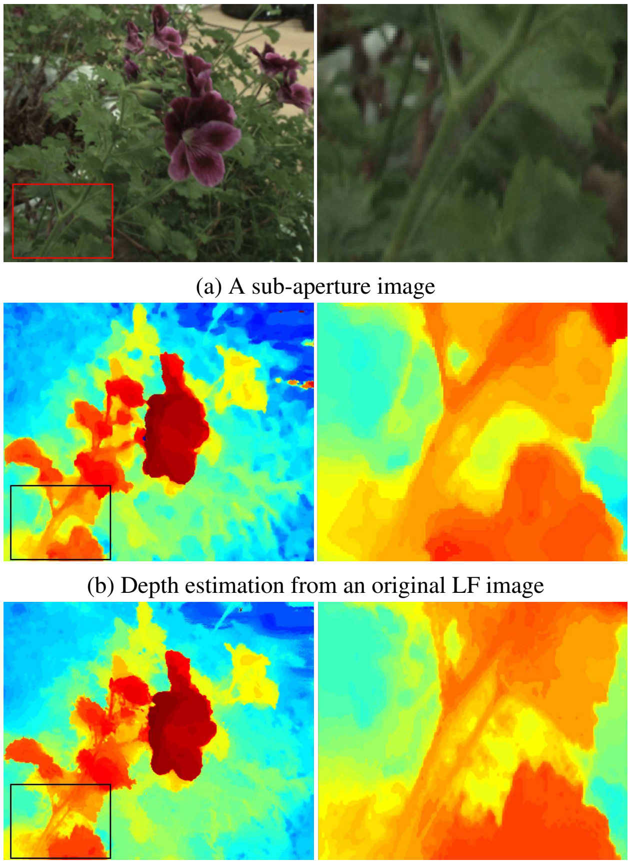
6.3. Applications

In this section, we demonstrate that the proposed method can enhance real-word LF imaging applications such as refocusing and depth estimation. Fig. 6 shows the refocusing results after applying different upsampling methods. We used the Light Field Toolbox [10] to generate refocusing results and tuned the user-controllable parameters of the toolbox to maximize the visual quality of the results. The focused regions of our method have sharper details than that of the other methods.

We also apply our method for depth estimation from a LF image. We used a stereo matching-based depth estimation algorithm [14] to find correspondences between sub-aperture images. As stated in [14], a LF image with high spatial and angular resolution is preferred to obtain accurate correspondences. In Fig. 7, we compare the estimated depth maps before and after applying our upsampling. The depth map from the upsampled image preserves fine details well as the high resolution image is more accurately discretized than the original image.

7. Conclusion

In this paper, we have presented the deep convolutional network for light-field image super-resolution. Our network has been concatenated with the spatial SR network and the



(a) A sub-aperture image

(b) Depth estimation from an original LF image

(c) Depth estimation after applying our upsampling method

Figure 7. Comparison of depth estimation before and after applying our upsampling.

angular SR network to enhance the spatial and angular resolution together. Therefore, our method can produce a high-resolution sub-aperture image in the novel view between adjacent sub-aperture views. We have demonstrated the effectiveness of our method through synthetic and real-world experiments compared to state-of-the-art LFSR methods. We also have applied our method to various applications such as image refocusing and depth estimation, and have shown promising results. In the future, we expect to extend our approach into temporal SR of light-field image sequences similar to [23].

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No.2010-0028680). Hae-Gon Jeon was partially supported by Global PH.D Fellowship Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2015H1A2A1034617).

References

[1] E. H. Adelson and J. R. Bergen. The plenoptic function and the elements of early vision. *Computational models of visual processing*, 1(2), 1991. 1

[2] F. Agostinelli, M. R. Anderson, and H. Lee. Adaptive multi-column deep neural networks with application to robust image denoising. In *Advances in Neural Information Processing Systems (NIPS)*, 2013. 2

[3] C. Birklbauer and O. Bimber. Panorama light-field imaging. *Computer Graphics Forum (CGF)*, 33(2):43–52, 2014. 1

[4] T. E. Bishop, S. Zanetti, and P. Favaro. Light field superresolution. In *IEEE International Conference on Computational Photography (ICCP)*, 2009. 2

[5] Y. Bok, H.-G. Jeon, and I. S. Kweon. Geometric calibration of micro-lens-based light-field cameras using line features. In *Proceedings of European Conference on Computer Vision (ECCV)*, 2014. 6

[6] D. Cho, S. Kim, and Y.-W. Tai. Consistent matting for light field images. In *Proceedings of European Conference on Computer Vision (ECCV)*, 2014. 1

[7] 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 *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2013. 2

[8] Z. Cui, H. Chang, S. Shan, B. Zhong, and X. Chen. Deep network cascade for image super-resolution. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 49–64. Springer, 2014. 1, 2, 5

[9] Y. L. Cun, B. Boser, J. S. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1:541–551, 1989. 1

[10] D. Dansereau. Ligh field toolbox v0.4. http://www.mathworks.com/matlabcentral/fileexchange/49683-light-field-toolbox-v0-4/. 7

[11] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *Proceedings of European Conference on Computer Vision (ECCV)*, 2014. 1, 2, 3, 4, 5

[12] D. Eigen, D. Krishnan, and R. Fergus. Restoring an image taken through a window covered with dirt or rain. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2013. 1, 2, 5

[13] S. Heber and T. Pock. Shape from light field meets robust pca. In *Proceedings of European Conference on Computer Vision (ECCV)*. 2014. 1

[14] H.-G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.-W. Tai, and I. S. Kweon. Accurate depth map estimation from a lenslet light field camera. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 3, 7

[15] Y. Jia. Caffe: Convolutional architecture for fast feature embedding. http://caffe.berkeleyvision.org/. 4

[16] K. I. Kim and Y. Kwon. Single-image super-resolution using sparse regression and natural image prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 32(6):1127–1133, 2010. 1

[17] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu. Saliency detection on light field. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 1

[18] C.-K. Liang and R. Ramamoorthi. A light transport framework for lenslet light field cameras. *ACM Transactions on Graphics*, 2015. 2

[19] Lytro. The lytro camera. http://www.lytro.com/. 1

[20] K. Mitra and A. Veeraraghavan. Light field denoising, light field superresolution and stereo camera based refocussing using a GMM light field patch prior. In *IEEE International Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),*, 2012. 2, 5, 6, 7

[21] Raytrix. 3d light field camera technology. http://www.raytrix.de/. 1

[22] C. J. Schuler, H. C. Burger, S. Harmeling, and B. Scholkopf. A machine learning approach for non-blind image deconvolution. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013. 2

[23] O. Shahar, A. Faktor, and M. Irani. Space-time super-resolution from a single video. In *CVPR*, 2011. 8

[24] L. Shi, H. Hassanieh, A. Davis, D. Katabi, and F. Durand. Light field reconstruction using sparsity in the continuous fourier domain. *ACM Transactions on Graphics*, 34(1):12, 2014. 2

[25] J. Sun, W. Cao, Z. Xu, and J. Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 1, 2, 5

[26] J. Sun, Z. Xu, and H.-Y. Shum. Image super-resolution using gradient profile prior. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2008. 1

[27] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi. Depth from combining defocus and correspondence using light-field cameras. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2013. 1, 7

[28] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing (TIP)*, 13(4):600–612, 2004. 5

[29] S. Wanner and B. Goldluecke. Spatial and angular variational super-resolution of 4d light fields. In *Proceedings of European Conference on Computer Vision (ECCV)*, 2012. 2, 3, 5, 6

[30] S. Wanner and B. Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 36(3):606–619, 2014. 2, 7

[31] S. Wanner, S. Meister, and B. Goldluecke. Datasets and benchmarks for densely sampled 4d light fields. In *Vision, Modelling and Visualization (VMV)*, 2013. 5

[32] L. Xu, J. S. Ren, C. Liu, and J. Jia. Deep convolutional neural network for image deconvolution. In *Advances in Neural Information Processing Systems (NIPS)*, 2014. 1, 2, 5

[33] Z. Yu, X. Guo, H. Ling, A. Lumsdaine, and J. Yu. Line assisted light field triangulation and stereo matching. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2013. 1