Combining Exemplar-based Approach and learning-based Approach for Light Field Super-resolution Using a Hybrid Imaging System

Haitian Zheng*[1], Minghao Guo†[2], Haoqian Wang‡[1], Yebin Liu§[2], and Lu Fang¶[1]

[1]Graduate School at Shenzhen, Tsinghua University, China
[2]Department of Automation, Tsinghua University, China

Abstract

We propose a new method to super-resolve images captured by a hybrid light field system that consists of a standard light field camera and a high-resolution standard camera. The high-resolution image is taken as a reference to help with super-resolving the low-resolution light field images. Our method combines an exemplar-based algorithm with the state of-the-art single image super-resolution approach and draws on the strengths of both. Both quantitative and qualitative experiments show that our proposed method substantially outperforms existing methods on standard light field datasets in the challenging large parallax setting.

1. Introduction

Light field imaging measures the spatial and angular variations in the intensity of light [3]. Because of the small volume and fast capturing speed of light field imaging systems, they have become one of the most extensively used devices for capturing a 3D scene. However, light field imaging systems has their own limitations. In their early years, light field cameras required expensive multi-camera arrays [4] to capture the high-resolution light-field image. In contrast, recent industrial light field cameras, such as Lytro [5] and RayTrix [6], are inexpensive and commercialized. However, suffering from restricted sensor resolution, light field cameras nowadays must make a trade-off between spatial and angular resolution [8].

In the literature, many studies have tried to break this trade-off, and they can be roughly categorized into three types. Firstly, some of the light field camera designs modify existing cameras and add programmable apertures [36] or coding modules [37] into the traditional camera design, and they use compressive sensing approach to recover high resolution spatial information. But this approach requires special hardware modification of the camera and is usually impractical. Secondly, [38, 39] apply single-image super-resolution methods to super-resolve the images captured by the light field camera. However, this method does not utilize the information from the reference image, and it usually does not perform well under large super-resolving scale (i.e. $\times 8$ scale). The last type of approach introduces a portable light field attachment which combines a traditional high-resolution digital SLR camera with low quality side cameras [7] or a light-field camera [8]. After applying exemplar-based image synthesis algorithms, low-resolution images are then transferred into high-resolution versions. This kind of approach only requires lightweight hardware modification, and it is able to synthesise high quality super-resolution output by combining information from the high-resolution reference image. As an exemplar-based approaches, however, it is not robust against the lack of exemplar patches. In other words, the output quality of such methods usually deteriorate severely under conditions such as occlusion or large parallax, resulting undesirable ghosting or blurring effects.

In this paper, we extend the work of [8] for light-field super-resolution using such hybrid imaging system because of its efficient hardware implementation and its desirable super-resolution output under non-occluded region. However, to overcome the ghosting and blurring effects that typically occurred under occlusion region or large parallax setting, we proposed a novel combined approach which take advantage of both the state-of-the-art single-image-based super-resolution approach [35] and the exemplar-based approach.

Specifically, unlike [8] which directly synthesises super-

*zheng.ht.ustc@gmail.com
†gmh14@mails.tsinghua.edu.cn
‡wangyizhai@sz.tisnghua.edu.cn
§liuyebin@mail.tsinghua.edu.cn
¶fanglu@sz.tsinghua.edu.cn

resolution output using the reference image, we instead synthesise an error map which measures the difference from the high-resolution ground-truth to the prediction of another independent algorithm (such as the single-image-based approach [35]). Our key insight is, the state-of-the-art single-image-based approaches provide robust low-frequency estimation, which usually perform better in the occluded region. Therefore, by applying exemplar-based approach upon such single-image-based estimation for better high-frequency detail enhancement, occlusion robust reconstruction can be further obtained. Extensive experiments validate that our proposed approach substantially outperform other methods, both quantitatively and qualitatively.

2. Related Work

Light fields [1, 11] provide a new angular dimension, allowing various visual applications such as light field display [12] and light field microscopy [13, 14]. Many recent works have tried to capture or synthesize high quality light fields from different types of input data. In this paper, we are interested in improving the spatial resolution of light field, using high-resolution light field images captured by different view to that of this high-resolution reference image.

2.1. Single Image Super-resolution

Single image SR resolution is a classic computer vision problem. To solve this problem, plenty of approaches have previously been proposed. These approaches typically utilize sparse representation [24, 25], neighbor embedding [26, 27], or anchored neighborhood regression [28, 29, 30]. Recently, deep-learning-based approaches have achieved state-of-the-art performance on single image super-resolution. Specifically, [31, 32] proposed a 3-layer convolutional neural network for image super-resolution (SRCNN). [35] further proposed a 20-layer convolutional neural network for the same task, which achieves even better performance. [33, 34] further discussed acceleration techniques for the deep learning based super-resolution approaches.

Although deep-learning-based approaches (such as VDSR) achieve impressive results, they fail to utilize the high frequency information, which is provided by the high-resolution reference image.

2.2. Light Field Super-resolution

In the studies of light field super-resolution, several methods have been proposed to restore the high frequency information using the provided high-resolution reference images. Specifically, [22] proposes a method to estimate both the high resolution depth map and luminous intensity in the Bayesian framework under the Lambertian textural prior. In addition, exemplar-based approaches, which have been widely used in texture synthesis [15], image completion [16], denoising [17] and deblurring [18], have also been applied to super-resolution [19, 20]. In light field research, [8] introduces a hybrid imaging system and a corresponding exemplar-based algorithm for super-resolution. [7] proposes a different light-field attachment and an iterative algorithm based on the work of [8]. However, exemplar-based approaches do not work well under the deficiency of exemplar samples, and occluded or specular highlighted regions usually cause the performance degeneration of such algorithms. As a consequence, although exemplar-based approaches such as [8] take advantages from additional high-resolution samples, the state-of-the-art single-image-based approaches still outperform such exemplar-based approaches by a large margin.

Other LF image/video interpolation methods are also related to our paper, in spite of the different application context. Specifically, [9] proposes a learning-based method for synthesizing new views from the sparsely sampled LF images. With an additional high frame-rate video, [10] aims to increase the temporal resolution of the low frame-rate LF video.

3. Proposed Method

This section introduces the proposed patch-based method which integrates the state-of-the-art single image SR algorithm, i.e. VDSR [35], for super-resolution of the side view images. The configuration is outlined in Figure 1. The basic concept is to use a patch-based method to fix the error made by the single image SR algorithm.

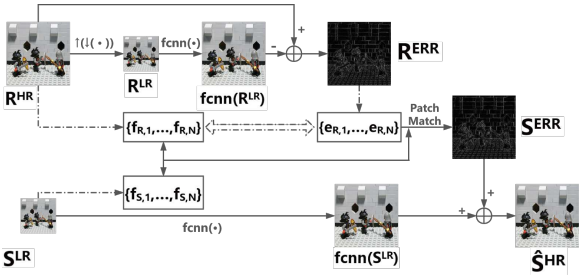

Figure 1: The overview of our proposed method. The HR reference image R^{HR} is firstly converted to the low-resolution version by $\uparrow(\downarrow(\cdot))$ and is then upsampled by the VDSR algorithm $f_{CNN}(\cdot)$ for computing the reference error map R^{ERR}. The source error map S^{ERR} is consequently estimated by applying patchmatch[8] on R^{ERR}. Finally, S^{ERR} is added to the VDSR-upsampled source image for producing the SR result.

We consider two input images: a low-resolution source image S^{LR}, and a high-resolution side-view reference image R^{HR}. The two images show the same scene in two different views.

3.1. Computing the Reference Error Map

Using the available high-resolution reference image, we can obtain a reference error map, which presents the error made by VDSR on the reference image. This reference error map can be of help for estimating the error made by VDSR on the source image. Specifically, let us denote the low-resolution version of the reference image as

$$R^{LR} = \uparrow (\downarrow (R^{HR})), \tag{1}$$

where $\downarrow (\cdot)$ is the downsampling operation, and $\uparrow (\cdot)$ is the bicubic upsampling operation. The super-resolved result of R^{LR} is computed by using the VDSR algorithm $f_{CNN}(\cdot)$. Afterward, the reference error map is calculated by

$$R^{ERR} = R^{HR} - f_{CNN}(R^{LR}). \tag{2}$$

3.2. Estimating Source Error Map

After obtaining the reference error map R^{ERR}, we adopt an exemplar-based method [8] for estimating the VDSR error map of the source image. Specifically, we extract patches from R^{HR}, and convert them into low-resolution patches for computing the gradient features dictionary $\mathbf{D}_R = \{f_{R,1}, ..., f_{R,N}\}$. Simultaneously, error patches at the same location of the reference error map R^{ERR} are also extracted, which is denoted as $\mathbf{E}_R = \{e_{R,1}, ..., e_{R,N}\}$.

To obtain the source error map, we calculate the feature f_j from the location j of the low-resolution source image S^{LR}. The 9 nearest neighbors in $\mathbf{D}_R$ with the smallest L_2 distance from f_j are computed. These 9 nearest neighbors in $\mathbf{D}_R$ (denoted as $\{f^j_{R,k}\}_{k=1}^{9}$) along with 9 error patches in $\mathbf{E}_R$ (denoted as $\{e^j_{R,k}\}_{k=1}^{9}$) are extracted. Then the reconstruction weights motivated from [8] are calculated. The estimated error patches $\hat{e_j}$ corresponding to location j are

$$\hat{e_j} = \frac{\sum_{k=1}^{9} w_k e^j_{R,k}}{\sum_{k=1}^{9} w_k}, w_k = exp\frac{-||f_j - f^j_{R,k}||^2}{2\sigma^2}. \tag{3}$$

Finally, the overlapping error patches $\hat{e_j}$ at every location j are averaged to estimate S^{ERR}, i.e., the VDSR error map of the source image.

3.3. Integrated Super-resolution

After obtaining the estimated error map S^{ERR}, which indicates the error of VDSR on the source image, the next natural step is to correct such error. Specifically, S^{ERR} is added to the VDSR super-resolved version of the source image:

$$\hat{S}^{HR} = f_{CNN}(S^{LR}) + S^{ERR}. \tag{4}$$

Here we explain the intuition behind such a scheme. In textural regions, a patch-based approach usually works better than a single-image approach as the reference image provides high-quality exemplar patches with accurate high-frequency details. However, in occluded or specular highlighted regions, a patch-based approach tend to fail because of the lack of exemplar patches. When exemplar-patch-based approach fails, single image approach like VDSR will provides a good high-resolution estimation. Our approach also resembles the subband decomposition, where an image is decomposed into a low-frequency single image approach band and a corresponding high-frequency residue. We apply VDSR to recover the low-frequency single image approach band, and apply an exemplar-based approach to recover the high-frequency band.

4. Experimental Results

We evaluate the performance of our proposed method for dense light field rendering on two dataset, namely the Stanford Light Field dataset [23] (from the Lego Gantry) and the Stanford Light Field dataset [23] (from the Gantry), which includes several challenging scenes with complex textures, specularity, and side view images with large parallax.

4.1. Experimental Setup

For both of the datasets, we select light field images of 2 views from the same horizontal line. The distance of the two light field images is set to 10, which is a considerably large parallax. The right images are down-sampled and regarded as low-resolution source images S^{LR}. The left images are regarded as the reference images R^{HR}. We evaluate our method in two different scales: $\times 4$ and $\times 8$.

We evaluate VDSR [35], a patch-based super-resolution algorithm [8] (denoted by PBSR in this section), and our approach on the two different scales. For VDSR [35] in the scale $\times 4$, we used the released model of [35] for super-resolution, while the $\times 8$ VDSR result is obtained by applying $\times 4$ VDSR upsampling followed by $\times 2$ bicubic upsampling. For PBSR [8] in scales $\times 4$ and $\times 8$, we set a patch size in low-resolution as 8×8, search range as 15 pixels, and $\frac{1}{2\sigma^2} = 0.0125$.

4.2. Super-resolution Results

Table 1(a) and Table1(b) show the performances of VDSR, PBSR and our approach on the Stanford dataset in the scales $\times 4$ and $\times 8$, respectively. It can be noticed that the PSNRs of our approach in both scales are higher than those of VDSR, since our approach takes advantages of the high-frequency detail provided by the reference image. Meanwhile, our approach outperforms PBSR. This is because PBSR is not specially designed to handle occlusion and specularity that usually occurs with large parallax, while our approach utilizes the VDSR for processing occlusions.

Figure 2 and Figure 3 illustrate super-resolution results on the dataset. By observing the zoomed-in results in row

Figure 2: Super-resolution result comparison in $\times 4$ scale. From top to bottom: (a) ground truth, (b) VDSR, (c) PBSR and (d) our method.

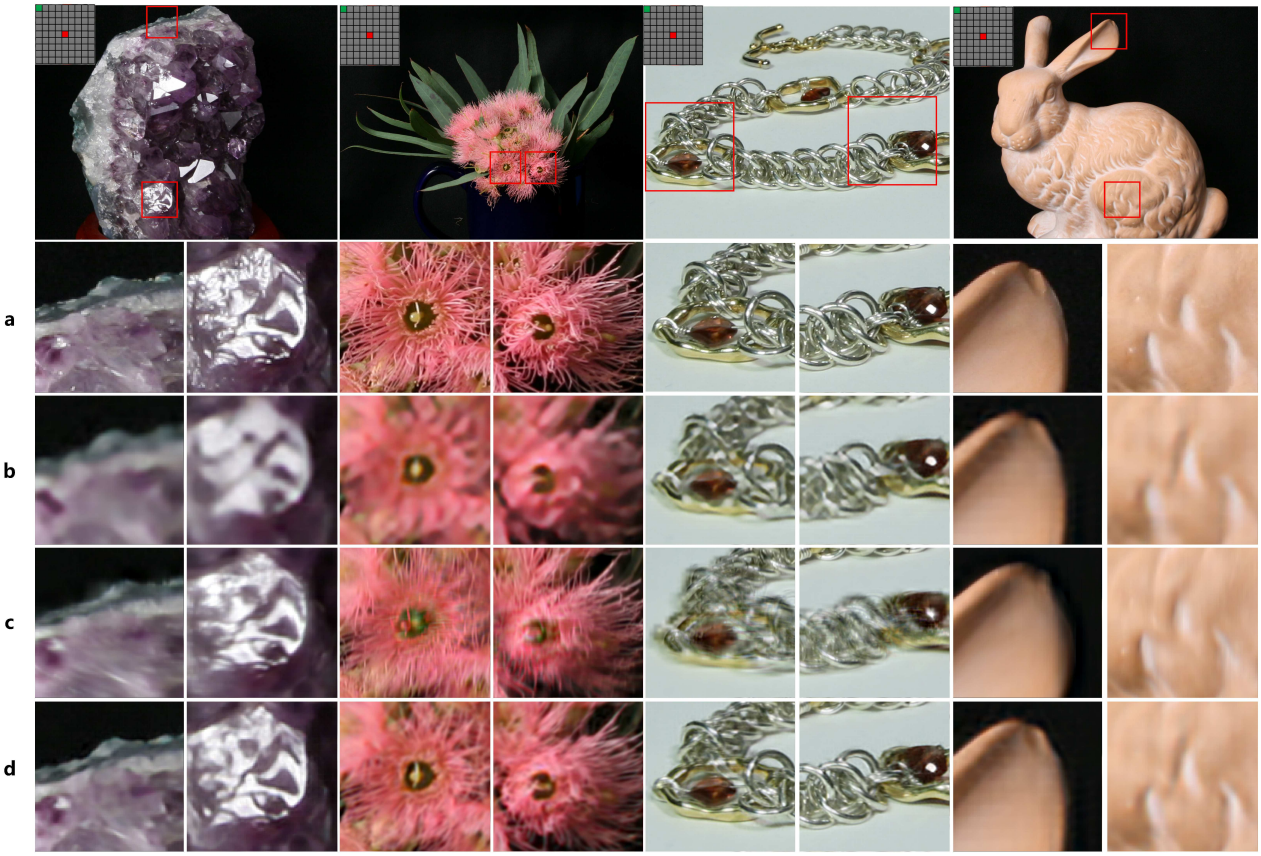

Figure 3: Super-resolution results comparison in $\times 8$ scale. From top to bottom: (a) ground truth, (b) VDSR, (c) PBSR and (d) our method.

(a) Scale $\times 4$ super-resolution results.

Image	Bicubic	VDSR [35]	PBSR [8]	ours
$Chess$	33.3402	35.5553	36.0067	**37.9788**
$LegoBulldozer$	31.479	34.6345	31.9685	**35.2196**
$LegoTruck$	33.8503	35.3746	37.1094	**38.0753**
$EucalyptusFlowers$	32.241	33.4495	33.0404	**33.8013**
$Amethyst$	32.6559	34.3501	34.8299	**35.7121**
$Bracelet$	28.8296	30.5893	29.2551	**31.1606**
$StanfordBunny$	39.1716	42.4671	40.9556	**43.2953**
$JellyBeans$	43.0806	45.5689	43.2104	**45.9222**
$LegoKnights$	30.9126	33.6056	31.5034	**34.0243**
$TarotCards(small)$	26.0773	28.0312	27.4733	**29.2274**
$TarotCards(large)$	25.0014	**27.4172**	24.6538	27.2570
$TreasureChest$	28.4255	29.6139	28.5932	**29.8761**
$LegoGantry$	28.8707	31.1338	31.2648	**32.544**
Average	31.8412	33.9840	33.0668	**34.9304**

(b) Scale $\times 8$ super-resolution results.

Image	Bicubic	VDSR [35]	PBSR [8]	ours
$Chess$	29.1794	30.7205	30.8882	**32.4780**
$LegoBulldozer$	27.048	29.261	27.173	**29.515**
$LegoTruck$	30.4047	31.4772	33.4653	**33.8647**
$EucalyptusFlowers$	29.8656	30.7956	30.4894	**30.9614**
$Amethyst$	29.0107	30.1847	31.6159	**32.0057**
$Bracelet$	24.5000	25.3199	25.1986	**26.0575**
$StanfordBunny$	33.6903	37.0836	35.7900	**38.3557**
$JellyBeans$	36.5531	40.4824	36.9044	**40.5603**
$LegoKnights$	26.9367	28.9750	27.2989	**29.2336**
$TarotCards(small)$	22.6408	23.9477	22.8769	**24.3882**
$TarotCards(large)$	21.6241	**22.6785**	21.1274	22.4126
$TreasureChest$	25.5891	26.2591	25.5121	**26.2956**
$LegoGantry$	24.8678	26.6092	27.7943	**28.9813**
Average	27.8393	29.5227	28.9334	**30.3931**

Table 1: $\times 4$ and $\times 8$ super-resolution experiment on the Stanford light field dataset [23] Light Fields from the Lego Gantry. PSNR comparison of bicubic interpolation, VDSR[35], patch-based method[8] and ours are listed in tables.

(d), it can be seen that our method has clearer high frequency details than those of VDSR in row(b) and PBSR in row(c).

Note that our algorithm was implemented with Matlab without parallel computing optimization. The run-time for the proposed algorithm is 3 minutes per image on an Intel i7 processor with 16 GB RAM.

5. Conclusion

In this work, we proposed an accurate super-resolution method for a hybrid imaging system. To improve the quality of the super-resolution result, we incorporated an exemplar-based approach with the state-of-the-art single image super-resolution algorithm. Experiments on the Stanford Light Field dataset demonstrate the substantial improvement of our approach both in quantity and quality.

(a) Scale $\times 4$ super-resolution results.

Image	Bicubic	VDSR [35]	PBSR [8]	ours
$CDs(occlued)$	27.3590	28.8787	27.3883	**29.9141**
$CDs(unocclued)$	27.8429	29.2925	29.0894	**29.7159**
$Humvee(occluded)$	42.6521	**43.4338**	41.5369	43.2132
$Humvee(unoccluded)$	29.9544	31.9358	31.8009	**33.0976**
Average	31.9521	33.3852	32.4539	**33.9852**

(b) Scale $\times 8$ super-resolution results.

Image	Bicubic	VDSR [35]	PBSR [8]	ours
$CDs(occlued)$	25.0226	26.2536	26.1992	**26.9569**
$CDs(unocclued)$	24.3284	25.3855	24.2227	**25.4434**
$Humvee(occluded)$	38.4171	**39.3918**	36.9326	38.9085
$Humvee(unoccluded)$	26.1588	27.6895	28.0290	**29.0252**
Average	28.4817	29.6801	28.8459	**30.0835**

Table 2: $\times 4$ and $\times 8$ super-resolution experiment on the Stanford light field dataset [23] Light Fields from the Gantry. PSNR comparison of bicubic interpolation, VDSR[35], patch-based method[8] and ours are listed in tables.

6. Acknowledgement

This work is supported in part by Natural Science Foundation of China (NSFC) under contract No. 61331015, 6152211, 61571259 and 61531014, in part by the National key foundation for exploring scientific instrument No.2013YQ140517, in part by Shenzhen Fundamental Research fund (JCYJ20170307153051701).

References

[1] Levoy M, Hanrahan P. "Light field rendering." SIGGRAPH ACM, 1996.

[2] Ihrke, Ivo, John Restrepo, and Lois Mignard-Debise. "Principles of Light Field Imaging: Briefly revisiting 25 years of research." IEEE Signal Processing Magazine 33.5 (2016): 59-69.

[3] Adelson E H, Bergen J R. "The plenoptic function and the elements of early vision." (1991).

[4] Wilburn B, et al. "High performance imaging using large camera arrays." ACM TOG. Vol. 24. No. 3. ACM, 2005.

[5] Lytro. https://www.lytro.com/.

[6] RayTrix.3D light field camera technology.http://lightfields.standford.edu/

[7] Wang Y, Liu Y, Heidrich W, et al. "The Light Field Attachment: Turning a DSLR into a Light Field Camera Using a Low Budget Camera Ring." IEEE TVCG (2016).

[8] Boominathan V, Mitra K, Veeraraghavan A. "Improving resolution and depth-of-field of light field cameras using a hybrid imaging system." ICCP IEEE, 2014.

[9] Kalantari N K, Wang T C, Ramamoorthi R. "Learning-based view synthesis for light field cameras." ACM TOG (Proc. of SIGGRAPH Asia 2016).

[10] Wang T C, Zhu J Y, et al, "Light Field Video Capture Using a Learning-Based Hybrid Imaging System." ACM TOG (Proc. of SIGGRAPH 2017).

[11] Gortler S J, Grzeszczuk R, Szeliski R, et al. "The lumigraph." SIGGRAPH ACM, 1996.

[12] Wetzstein G, Lanman D, Hirsch M, et al. "Compressive light field displays." IEEE computer graphics and applications 32.5 (2012): 6-11.

[13] Levoy M, Ng R, Adams A, et al. "Light field microscopy." ACM TOG 25.3 (2006): 924-934.

[14] Prevedel R, Yoon Y G, Hoffmann M, et al. "Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy." Nature methods 11.7 (2014): 727-730.

[15] Efros A A, Freeman W T. "Image quilting for texture synthesis and transfer." SIGGRAPH ACM, 2001.

[16] Sun J, Yuan L, Jia J, et al. "Image completion with structure propagation." ACM TOG 24.3 (2005): 861-868.

[17] Buades A, Coll B, Morel J M. "A non-local algorithm for image denoising." CVPR 2005.

[18] Cho S, Wang J, Lee S. "Video deblurring for hand-held cameras using patch-based synthesis." ACM TOG 31.4 (2012): 64.

[19] Freedman G, Fattal R. "Image and video upscaling from local self-examples." ACM TOG 30.2 (2011): 12.

[20] Freeman W T, Jones T R, Pasztor E C. "Example-based super-resolution." IEEE Computer graphics and Applications 22.2 (2002): 56-65.

[21] Georgiev T. "New results on the Plenoptic 2.0 camera." Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on. IEEE, 2009.

[22] Bishop, Tom E., Sara Zanetti, and Paolo Favaro. "Light field superresolution." ICCP, 2009.

[23] Standard,"The (new) standford light field archive,http://lightfield.standford.edu,2008

[24] Yang J, Wright J, Huang T, et al. "Image superresolution as sparse representation of raw image patches." CVPR, 2008.

[25] Yang J, Wright J, Huang T S, et al. "Image superresolution via sparse representation." ICIP, 2010.

[26] Chang H, Yeung D Y, Xiong Y. "Super-resolution through neighbor embedding." CVPR, 2004.

[27] Bevilacqua M, Roumy A, Guillemot C, et al. "Low-complexity single-image super-resolution based on nonnegative neighbor embedding." (2012): 135-1.

[28] Yang C Y, Yang M H. "Fast direct super-resolution by simple functions." ICCV, 2013.

[29] Timofte R, De Smet V, Van Gool L. "Anchored neighborhood regression for fast example-based superresolution." ICCV, 2013.

[30] Timofte R, De Smet V, Van Gool L. "A+: Adjusted anchored neighborhood regression for fast superresolution." ACCV, 2014.

[31] Dong C, Loy C C, He K, et al. "Learning a deep convolutional network for image super-resolution." ECCV, 2014.

[32] Dong C, Loy C C, He K, et al. "Image super-resolution using deep convolutional networks." IEEE PAMI 38.2 (2016): 295-307.

[33] Shi W, Caballero J, Huszr F, et al. "Real-time single image and video super-resolution using an efficient subpixel convolutional neural network." CVPR, 2016.

[34] Dong C, Loy C C, Tang X."Accelerating the superresolution convolutional neural network." ECCV,2016.

[35] Kim J, Kwon Lee J, Mu Lee K. "Accurate image superresolution using very deep convolutional networks." CVPR, 2016.

[36] Liang C K, Lin T H, Wong B Y, et al. "Programmable aperture photography: multiplexed light field acquisition." ACM TOG Vol. 27. No. 3. ACM, 2008.

[37] Marwah K, Wetzstein G, Bando Y, et al. "Compressive light field photography using overcomplete dictionaries and optimized projections." ACM TOG 32.4 (2013): 46.

[38] Pujades S, Devernay F, Goldluecke B. "Bayesian view synthesis and image-based rendering principles." CVPR 2014.

[39] Shi L, Hassanieh H, Davis A, et al. "Light field reconstruction using sparsity in the continuous fourier domain." ACM TOG 34.1 (2014): 12.