Cross-Scale Reference-Based Light Field Super-Resolution

Mandan Zhao, Gaochang Wu, Yipeng Li, Xiangyang Hao, Lu Fang, *Senior Member, IEEE*, and Yebin Liu, *Member, IEEE*

***Abstract*—Light fields suffer from a fundamental resolution tradeoff between the angular and the spatial domain. In this paper, we present a novel cross-scale light field super-resolution approach (up to 8× resolution gap) to super-resolve low-resolution (LR) light field images that are arranged around a high-resolution (HR) reference image. To bridge the enormous resolution gap between the cross-scale inputs, we introduce an intermediate view denoted as single image super-resolution (SISR) image, i.e., super-resolving LR input via single image based super-resolution scheme, which owns identical resolution as HR image yet lacks high-frequency details that SISR scheme cannot recover under such significant resolution gap. By treating the intermediate SISR image as the low-frequency part of our desired HR image, the remaining issue of recovering high-frequency components can be effectively solved by the proposed high-frequency compensation super-resolution (HCSR) method. Essentially, HCSR works by transferring as much as possible the high-frequency details from the HR reference view to the LR light field image views. Moreover, to solve the nontrivial warping problem that induced by the significant resolution gaps between the cross-scale inputs, we compute multiple disparity maps from the reference image to all the LR light field images, followed by a blending strategy to fuse for a refined disparity map; finally, a high-quality super-resolved light field can be obtained. The superiority of our proposed HCSR method is validated on extensive datasets including synthetic, real-world and challenging microscope scenes.**

***Index Terms*—Light field, super-resolution, reference-based super-resolution, depth estimation.**

I. Introduction

A LIGHT field can be defined as the collection of all light rays in a 3D space [1]–[3]. Unlike regular cameras, plenoptic (light field) cameras [4] capture directional light information, which generates new capabilities including adjustment of the camera parameters (such as focus and aperture size), changes in the camera viewpoint, and estimation of depth. As a result, light field imaging is increasingly being used in a variety of application areas including digital photography, microscopy, robotics, and computer vision. Conventional light field capture systems, such as multi-camera arrays [5] and light field gantries [6], require expensive custom-made hardware. In recent years, commercially available light field cameras have been developed, such as the Lytro [7] and RayTrix [8], which include a micro-lens array as well as have the capacity for simultaneous capture. Unfortunately, due to the restricted sensor resolution, they usually suffer from a resolution trade-off between the spatial and angular domains. So improving the resolution of the light field has become a hot-spot of these related areas.

To improve the spatial resolution of the light field while maintaining the angular resolution, the hybrid camera setup [11] using a Lytro camera and a high-resolution DSLR camera is proposed. Using a PatchMatch-based super-resolution (PaSR) method [11], some of the high-frequency details of the DSLR camera can be transferred to the dense Lytro views to give a spatial resolution-improved light field. However, PaSR computes the target pixel value as the average value of all the overlapping patches on the DSLR, which leads to a loss of high-frequency information. An iterative Patch-And-Depth-based Synthesis (iPADS) method [12] is further proposed by deforming the candidate patches based on the surface geometry, such that the averaged patches are more similar to the ground-truth to mitigate the high-frequency loss in the average operation. However, the synthesized super-resolved images still suffer from inaccuracy due to the distortion of the disparity maps and the fusion operation of multiple patches, especially on the occlusion regions, as shown in Fig. 1(b) and (c).

In this paper, we discard the conventional ways that apply patch-based averaging, which will blur the high-frequency details naturally. To bridge the enormous resolution gap between the cross-scale inputs (usually up to 8×), we introduce an intermediate view denoted as SISR image: i.e., super-resolving LR input via Single Image based Super-Resolution scheme (here we use VDSR [10] followed by bicubic upsampling to super-resolve LR image by 8×), serving as the 'middleman'. The insight lies in that the SISR image owns identical resolution as HR, while it lacks high frequency details that can hardly be recovered by SISR scheme directly, especially for such signifi-

Manuscript received October 17, 2017; revised March 20, 2018; accepted April 30, 2018. Date of publication May 24, 2018; date of current version August 13, 2018. This work was supported in part by the National Key Foundation for Exploring Scientific Instrument under Grant 2013YQ140517 and in part by the National Natural Science Foundation of China under Grants 61522111, 61531014, 61722209, and 61331015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Oliver Cossairt. *(Corresponding authors: Lu Fang and Yebin Liu.)*

M. Zhao and X. Hao are with the Zhengzhou Institute of Surveying and Mapping, Zhengzhou 450001, China (e-mail: mandanzhao@163.com; xiangyanghao2004@163.com).

G. Wu is with the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110004, China (e-mail: 790723977@qq.com).

Y. Li and Y. Liu are with the Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: liuyebin@tsinghua.edu.cn; liuyebin@mail.tsinghua.edu.cn).

L. Fang is with the Center for Data Science and Information Technology, Tsinghua-Berkeley Shenzhen Institute, Shenzhen 518055, China (e-mail: fanglu@sz.tsinghua.edu.cn).

Digital Object Identifier 10.1109/TCI.2018.2838457

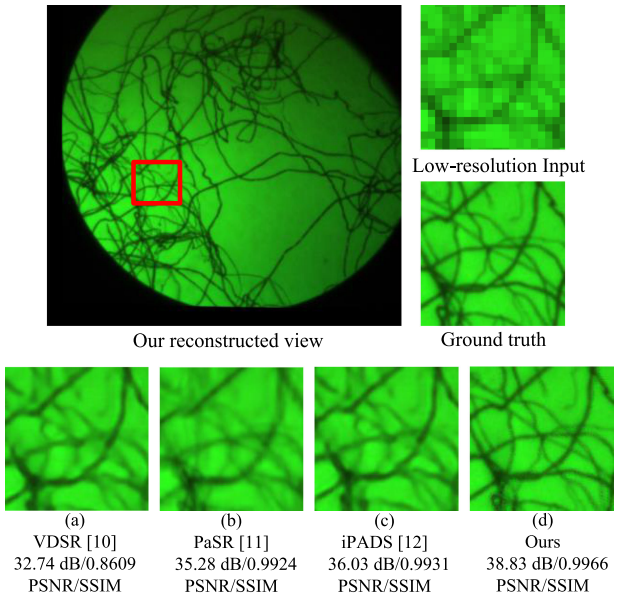


Fig. 1. Comparison of light field super-resolved results by a factor of 8 on microscope light field data *Cotton* [9] obtained by the VDSR method [10], PaSR method [11], iPADS method [12] and our proposed approach. The result of the VDSR [10] method shows a serious blur in the entire scene because of the large super-resolution scale. The result of the PaSR [11] method shows ghosting artifacts, especially in the occlusion regions. The result of the iPADS method [12] suffers from distortion because of the inaccuracy of the disparity maps. The proposed HCSR method produces better subjective results in this challenging case, achieving the highest PSNR and SSIM values (a) VDSR [10] 32.74 dB/0.8609 PSNR/SSIM, (b) PaSR [11] 35.28 dB/0.9924 PSNR/SSIM, (c) iPADS [12] 36.03 dB/0.9931 PSNR/SSIM, (d) Ours 38.83 dB/0.9966 PSNR/SSIM Low-resolution Input.

cant resolution gap. Interestingly, we can treat the intermediate SISR image as the low frequency part of our desired HR image, then the remaining issue would be how to recover the high frequency components? Along this line, we propose a novel high-frequency compensation super-resolution (HCSR) method to transfer as much as possible the high-frequency details from the high-resolution reference view to the low-resolution light field image views. We consider a similar input – several low-resolution side views arranged around a central high-resolution view – as [12], where the central view serves as a reference for the super-resolution of all the side views. The SISR image for the central view is used for the extraction of the high-frequency components that are missing in the side views, while the SISR image of the side views are used for disparity computation and as the basis for adding the high-frequency information wrapped from the central view based on the computed disparity map. In the disparity estimation step, we propose a multi-disparity blending strategy, which takes advantage of the regular layout structure of the light field image array and applies hole-filling procedures to obtain the final refined disparity map. In this way, we can avoid the blurry effect originated from the patch averaging operations in [11] and [12], so as to optimize the high frequency perseverance of reconstruction. The main features and contributions of our proposed HCSR are elaborated as follows:

- **Cross-scale**: We introduce the intermediate SISR view for both of the high-resolution and low-resolution inputs to bridge the significant resolution gap, so that we can handle the large scaling factor ($\times 8$).
- **Large parallax**: Parallax introduces occlusion regions, which is the main challenge in available warping problems. We utilize complementary information of the occlusion regions from different side views of the light field and propose a fusion and hole-filling algorithm to compute a high-quality depth for the reference view. The maximum disparity between the reference image and the side view images is about 35 pixels counted on the reference image.
- **Retaining details**: Extensive experiments validate that the proposed approach significantly improves the reconstruction quality subjectively and objectively (PSNR: average 2 dB higher) compared to other state-of-the-art approaches.
- **Low complexity**: In addition, our framework is significantly faster than the PatchMatch-based method [11], [12] for reconstruction of the whole light field (more than 20 times faster). This shows the efficiency of our approach, validating the practical usage of the proposed algorithm. In addition, we demonstrate the applications of our approach for depth enhancement using the reconstructed high spatial resolution light field.

The source code of this work will be made public. The rest of the paper is organized as follows. In Section II we discuss the related work. Section III illustrates the proposed HCSR method for light field super-resolution. In Section IV, our detailed experimental results and analysis are provided. Section V draws a conclusion.

II. Related Work

A comprehensive overview of light field techniques is provided in Wu *et al.* [3]. In this paper, we mainly focus on the literature for improving the spatial resolution of the light field. We divide the related works into three categories: single image super-resolution, light field super-resolution and hybrid imaging-based super-resolution.

A. Single Image Super-Resolution

SISR is a classical computer vision problem that has been intensively studied. A classical SISR method can be categorized into four types: prediction models, edge-based methods, image statistical methods and example-based methods. Further details for the evaluation of these approaches can be found in the work of Yang *et al.* [13]. Among these, the example-based methods achieved state-of-the-art performance.

Recently, deep learning-based approaches achieved better performance on SISR. Dong *et al.* [14] proposed a network for SISR named SRCNN, in which a high-resolution image is predicted from a given low-resolution image. Kim *et al.* [10] improved on this work by using a residual network with a deeper structure and named the new method VDSR (Very Deep convolutional network for Super-Resolution). However, these SISR approaches cannot handle a scaling factor larger than 4 times. The capacity for restoring high-frequency information weakens quickly with the increasing of the scaling factor, appearing as an extremely blurry result in the super-resolved image.

B. Light Field Super-Resolution

Since a light field has limited resolution, many methods have been proposed to increase its spatial or angular resolution. Wanner and Goldluecke [15] introduced a variational light field spatial and angular super-resolution framework. Given the depth estimates at the input views, they reconstructed novel views by minimizing an objective function that maximizes the quality of the final results. Based on Wanner and Goldluecke's work, a certainty map was proposed to enforce visibility constrains on the initial estimated depth map in [16]. Yoon *et al.* [17] used convolutional neural networks (CNNs) to perform spatial and angular super-resolution. However, these methods could usually only increase the resolution by a scaling factor of 2 or 4. Zhang *et al.* [18] proposed a phase-based approach to reconstruct light fields. However, their method was designed for a micro-baseline stereo pair. Kalantari *et al.* [19] used two sequential CNNs to model depth and color estimation simultaneously by minimizing the error between synthetic views and ground truth images. Wu *et al.* [20] presented a CNN-based approach on the epipolar plane image (EPI) domain to reconstruct missing views. However, both of these two methods only increased the angular resolution of the light field.

In general, light field super-resolution contains spatial and angular domains. Some methods [18]–[20] only increase the resolution of one domain, and others [15]–[17] increase them both. However, the increase of the light field resolution is extremely limited. Meanwhile, the super-resolved results may have many artifacts because it is very difficult for most super-resolution algorithms to reconstruct high-frequency details from completely unknown information. Therefore, we need auxiliary information to form a hybrid input for better reconstruction of light fields in a larger scaling factor.

C. Super-Resolution Using Hybrid Input

The idea of hybrid imaging was proposed in the context of motion deblurring [21], where a low-resolution high-speed video camera co-located with a high-resolution still camera was used to deblur the blurred images. Following this, several examples of hybrid imaging have been found in different applications. Cao *et al.* [22] proposed a hybrid imaging system consisting of a RBG video camera and a low-resolution multi-spectral camera to super-resolve the high-resolution single spectral camera. Another example of a hybrid imaging system is the virtual view synthesis system proposed by Tola *et al.* [23], where four regular video cameras and a time-of-flight sensor are used. They show that by adding the time-of-flight camera, they could render better quality virtual views than by just using a camera array with similar sparsity. Recently, a high-resolution camera, co-located with a Shack-Hartmann sensor was used to improve the resolution of 3D images from a microscope [24]. Most of the above-mentioned hybrid imaging systems require a beam splitter to guarantee that the two streams are imaged with the same optical center without the need of disparity computation.

Boominathan *et al.* [11] introduced a PatchMatch-based light field super-resolution method using a hybrid imaging technique, where the patches from each view of a light field are matched with a reference high-resolution image of the same scene. Since the high-resolution image has the exact details of the scene, the super-resolved light field has the true information compared to the hallucinated information by [25], [26]. However, they use the library for approximate nearest neighbors [27] to search for matching patches in the reference high-resolution image, so the algorithm will be too slow to generate the whole super-resolved light field. Since the Lytro camera can only capture burst images with maximum frame rate at 1fps, the recent work of Wang *et al.* [28] generates a full light field video at 30 fps using a learning-based hybrid imaging based on the setup proposed in [11]. In this work, they only interpolated in the time dimension without super-resolve in the spatial dimension.

Wang *et al.* [12] proposed the design of a central-view camera along with a set of low-resolution side-view cameras for high quality light field acquisition. Compared with the setup in [11] using a combination of Lytro and DSLR, each side-view camera in this setup has independent optical design, which can be more flexible when choosing the size of the baseline, or other system parameters like focal length, spatial temporal resolution, etc. More importantly, it allows for the capture of light field video (Lytro cannot capture video). In all, compared with the setups which use cameras of the same resolution, the hardware setup in [12] saves the cost, minimizes the size of system, and also minimizes the data amount of the input (each side-view camera only takes about 1.56% data rate of the central camera with $8\times$ scale difference). Based on this hybrid setup, the authors proposed the iPADS method which combines PatchMatch-based and depth-based synthesis iteratively to update the patch database. Although the database is enriched with better patches, they finally synthesized the results using the PatchMatch approach, i.e., using multiple patches for averaging to get the final results, which inherently blurs the high frequency details. Comparably, in this paper, we discard these patch-based averaging approaches [11], [12], and proposed a new method that can separate the high frequency components in the input high-resolution image and transfer them to the desired image. In this way, high frequency texture can be reserved and better super-resolution results can be achieved.

III. Proposed High-Frequency Compensation Based Super Resolution (HCSR) Method

In this section, we introduce the proposed HCSR method. Note that the input to our framework is hybrid and cross-scale $n \times n$ light field; i.e., it contains $n \times n - 1$ low-resolution side views $\mathcal{L}_i, (i = 1, ..., N, N = n^2 - 1)$ arranged around a central high-resolution reference view $\mathcal{R}$. Our goal is to super-resolve all the $\mathcal{L}_i$ views using reference $\mathcal{R}$.

A. Overview of HCSR

Fig. 2 illustrates the pipeline of our proposed HCSR, which is composed of two main modules: the computation of the low frequency component of desired HR image: i.e., the intermediate SISR image of both $\mathcal{R}$ and $\mathcal{L}_i$ views (Module 1), as well as the extraction of the high-frequency component of desired HR image: i.e., the warping-based high-frequency com-

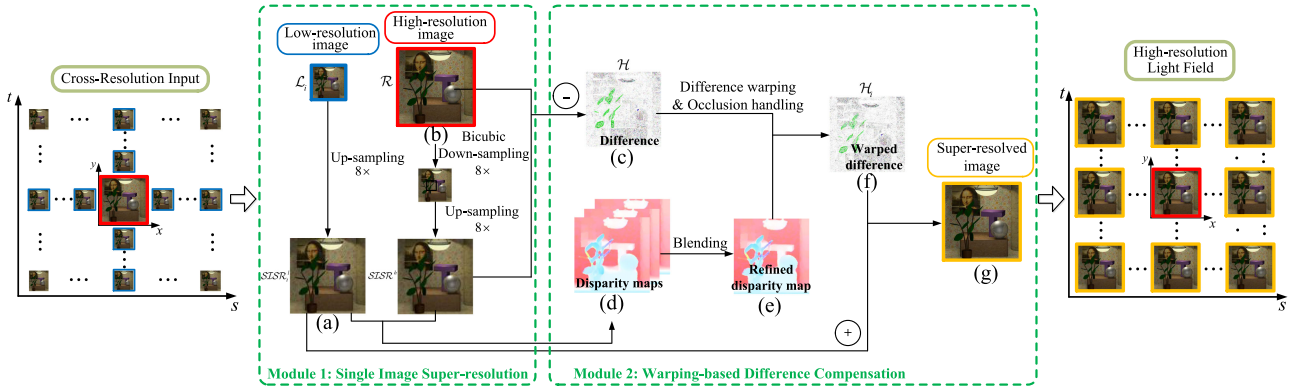


Fig. 2. Overview of the pipeline of the proposed HCSR approach for reconstructing light field.

pensation aiming to super-resolve each side view $\mathcal{L}_i$ (Module 2).

In Module 1, in order to compute the dense warping field to warp the high-frequency components of the central image, the enormous resolution gap between $\mathcal{R}$ and $\mathcal{L}_i$ views should be first bridged. So, the SISR views are proposed to balance the large resolution gap. Moreover, the computed SISR views are the key to obtaining the high-frequency components. Briefly, the SISR images $\mathcal{SISR}_i^l$ of $\mathcal{L}_i$ are generated using the available single image super-resolution (SISR) method (here we use VDSR [10] followed by bicubic upsampling to super-resolve LR image by $8\times$), while the SISR images $\mathcal{SISR}^h$ of $\mathcal{R}$ is computed by first downsampling it to the same resolution as the $\mathcal{L}_i$, followed by upsampling via the same method applied on $\mathcal{L}_i$ views.

In Module 2, we first compute the high-frequency details of the central view as the difference map (Fig. 2(c)) between $\mathcal{R}$ and its SISR image $\mathcal{SISR}^h$. In parallel, we compute the disparity maps (Fig. 2(d)) between each $\mathcal{SISR}_i^l$ and $\mathcal{SISR}^h$ and fuse them together to obtain a refined disparity map (Fig. 2(e)). Later, based on this refined disparity map, we propagate the difference map to each of the side views to form the warped disparity maps (Fig. 2(f)). Finally, the warped difference map is added back to the SISR images $\mathcal{SISR}_i^l$ to get the super-resolved images (Fig. 2(g)) of the whole light field.

B. Computing the Difference Map

To get the SISR view of each side view image $\mathcal{L}_i$, we up-sampled it to the same resolution of $\mathcal{R}$ using one of the available SISR method VDSR [10]. Note that state-of-the-art single image super-resolution (SISR) methods such as VDSR [10] have a decent super-resolution performance in the case of an up-sampling factor around $\times 2 - \times 4$. However, once the up-sampling goes to a large scale such as $\times 8$, the quality of SISR results degrades significantly, and most of the high-frequency details are not recovered; as shown in Fig. 1(a), the obtained SISR images $\mathcal{SISR}_i^l, (i = 1, ..., N)$ are blurred.

In order to recover the high-frequency details, our key idea is to transfer the high-frequency information contained in the central view image $\mathcal{R}$ to each input low-resolution side-view image. We calculate this high-frequency information as a difference map by:

$$\mathcal{H} = \mathcal{R} - \mathcal{SISR}^h = \mathcal{R} - (\mathcal{R}\downarrow)\uparrow. \tag{1}$$

Here, the difference map is obtained by the difference between $\mathcal{R}$ and its SISR version, $\mathcal{SISR}^h$, which is defined first by the bicubic down-sampling operation $\downarrow$ to match the size of $\mathcal{SISR}_i^l$, followed by the upsampling operation $\uparrow$ using the SISR method.

C. Computing the Disparity Maps

To warp the difference map $\mathcal{H}$ to all the side views, the disparity map of the central view, which stands for the dense correspondences between the central image to any of the side views, needs to be computed. Here, it is more reasonable to use both of the SISR images of the central view and the side views for disparity map computing, since they contain the same level of detail.

Available light field depth estimation methods [29], [30] can be adopted to calculate the disparity map in our case. However, most of these available methods assume that the light fields are captured perfectly with epipolar lines both horizontally and vertically well aligned, such as light fields captured from Lytro cameras. However, for light fields captured by camera arrays [5], these assumptions are difficult to hold and the performances of these methods are not guaranteed. We therefore use a more robust disparity calculation.

We calculate the disparity map as a multiple disparity fusion strategy. For simplicity, we only use the side views on the central cross-pattern of the light field.[1] We first utilize binocular stereo matching [31] to get $2 \times (n-1)$ versions of the disparity maps between the central SISR image $\mathcal{SISR}^h$ and all the other $\mathcal{SISR}_i^l$ images on the cross of the light field structure. Fig. 3(a) shows a light field structure with $n = 5$. In this case, 8 disparity maps of the central view are obtained. Unfortunately, these disparity maps are also usually inaccurate, in particular in the occlusion regions.

Because of the horizontal and vertical symmetry of the light field, the corresponding regions of each image pair (e.g., $\mathcal{L}_1$

[1]We note that occluded points would be visible from the diagonal views but not horizontal or vertical views. However, this rarely happens. To save the computation cost, we use the cross-pattern views for disparity calculation merely.

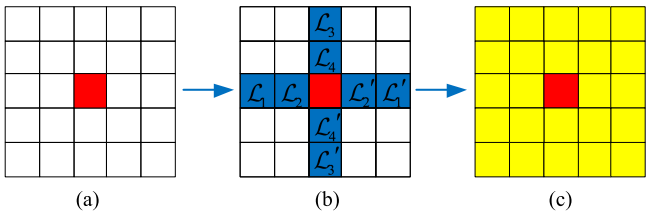


Fig. 3. The process of calculating the disparity map with our cross-pattern strategy. The red box in the central view is the $\mathcal{H}_i$ image. (b) The process of the stereo matching between the $\mathcal{SISR}_h$ image and $\mathcal{SISR}_l$ image (blue boxes) in the horizontal and vertical coordinate. (c) We generate the refined disparity maps (yellow boxes) of the whole light field.

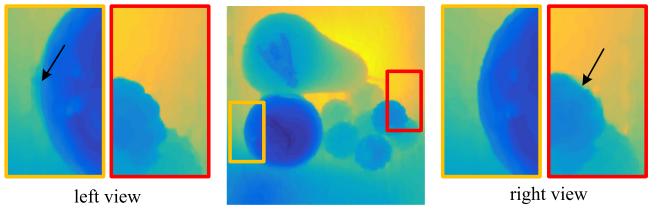


Fig. 4. The disparity map before and after replacement. The arrow pointing regions in close-up versions is not clear because of the occlusion. Thus these regions can be replaced by the values from the regions of the corresponding complementary disparity map.

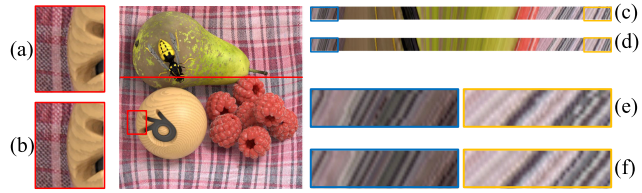


Fig. 5. Comparison of the super-resolved results and their EPI determining whether to adopt our cross-pattern strategy, including information complementarity and blending. (a) and (b) are the close-up versions of the super-resolved results in the red boxes. (c) and (d) are the EPIs located at the red line shown in the central view, which are upsampled to an appropriate scale for better viewing. (e) and (f) show the close-up versions of the EPIs (c) and (d) in the blue and yellow boxes, respectively. (a) and (c) are the results without using our proposed strategy. (b) and (d) are the results using our strategy, which show reasonable results with a smooth edge and continuous structure of the EPI.

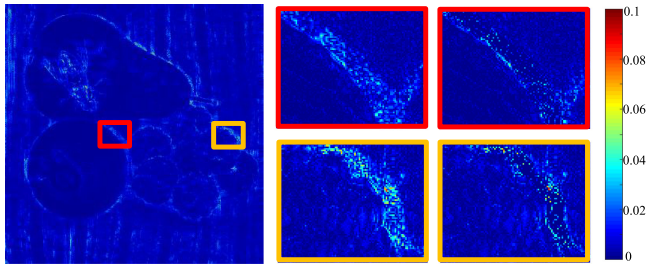


Fig. 6. (a) The error map between the super-resolved result and the ground truth in the *StillLife* dataset. The large red and yellow boxes show the close-ups of the red and yellow portions of this error map, respectively. (b) The results that do not use our hole-filling method on the occlusion regions. (c) The results that use our hole-filling method.

and $\mathcal{L}_1{}'$ in Fig. 3(b)) contain complementary information in the occlusion regions. We therefore divide the side view images into $n-1$ pairs. For example, as shown in Fig. 3(b)), $\mathcal{L}_1$ and $\mathcal{L}_1{}'$ are in a pair, and $\mathcal{L}_2$ and $\mathcal{L}_2{}'$ are in a pair.

For each side view image, we first detect the occluded regions on the central view that are not visible on the side view according to the left-right consistency check [32]. For example, we can mark the occluded regions on view $\mathcal{L}_1$ and view $\mathcal{L}_1'$. Here, to ensure that the real occlusion regions are included, we apply a dilation operation on the detected regions. For each occluded region detected by the disparity map $\mathcal{D}_i$, the disparity values can be replaced by the values from the regions of the corresponding complementary disparity map $\mathcal{D}_i'$, and vice versa. Such replacement is applied on all the disparity pairs and the accuracy of the $2 \times (n-1)$ disparity maps are improved. The refined disparity map is shown in the middle of Fig. 4. In the occlusion regions (the yellow and red boxes in Fig. 4), the edges are clear. In particular, the close-up versions in left and right views reflect this corresponding complementary strategy. The regions in the arrow pointing are bad and can be replaced by the values from the regions of the corresponding complementary disparity map. Fig. 4 shows the disparity map pair before and after replacement.

Finally, we blend all the improved disparity maps of the central view to form a refined disparity map (Fig. 2(e)). For each disparity value d_k of a pixel p on disparity map $\mathcal{D}_k$, $(k = 1, ..., 2(n-1))$, we discard the two maximum values and two minimum values to remove outliers. We then compute the mean value of the residue values to form the refined disparity map (Fig. 2(e)).

The performance of the proposed disparity blending scheme is validated in Fig. 5. It compares the final super-resolved light field (after the warping step) on the EPI, with and without the proposed blending scheme. Without the blending scheme, the disparity map of each side view is computed independently and the final high-frequency component warping result contains blurring and tearing artifacts in the occluded regions Fig. 5(a). The disparity blending strategy produces plausible results in the occluded regions (see Fig. 5(b)). Fig. 5(e) and (f) show the EPIs with and without our blending scheme. From the clearly structured EPI results (Fig. 5(f)), which show continuity (i.e., consistency among views), we can see that our reconstructed light field benefits from these high-quality and view-consistent disparities.

D. The Difference Map Warping and Occlusion Handling

In this subsection, we propagate the central difference map $\mathcal{H}$ to all the side views based on the refined disparity map. Because the disparity values are discrete, direct warping of the difference map produces quantized errors. To this end, we operate this warping step in the sub-pixel level by enlarging the size of the disparity map and difference map by a factor of 4 and then warp each pixel value to the side views based on the flow values. We finally down-sample the warped results to the original size.

Because of occlusions, we note that such a per-pixel based warping strategy produces unwarpped pixels that appear like the punctiform area on the side views as shown in Fig. 6(b). To mitigate these artifacts, we fill these holes (unwarped pixels)

through a bilateral filtering operation [33], which is based on the assumption that pixels with similar colors around a certain region are likely to have a similar value. We therefore utilize the pixel difference information of the holes' surroundings and estimate the difference in the occlusion regions by combining the value of color similarity and Euclidean distance. We assume that $\mathcal{W}$ is the square window of length γ ($\gamma = 7$) centered at the pixel (i, j). $g(i, j)$ is the difference value in these holes region and determined by all pixels in this window$\mathcal{W}$:

$$g(i,j) = \frac{\sum_{k,l\in\mathcal{W}} f(k,l)\omega(i,j,k,l)}{\sum_{k,l\in\mathcal{W}} \omega(i,j,k,l)}, \tag{2}$$

where i, j are the indices of the current unwarped pixel in the difference map, namely, the coordinate value. k, l are the indices of the neighbor pixel in that window $\mathcal{W}$. All the pixels are in the same coordinate system. $f(k, l)$ is a neighbor pixel difference value, and ω is the weight coefficient that is determined by the two constraints about the color similarity and their distance, i.e.,

$$\omega = c \times d, \tag{3}$$

where

$$c(i,j,k,l) = \exp\left(-\frac{\|f(i,j) - f(k,l)\|^2}{\gamma_c}\right), \tag{4}$$

and

$$d(i,j,k,l) = \exp\left(-\frac{(i-k)^2 + (j-l)^2}{\gamma_d}\right). \tag{5}$$

Here, c is the color similarity constraint, which can better keep the texture similarity, and d is the distance constraint, which means the closer the distance, the higher the similarity. γ_c and γ_d are two constants used as the thresholds of the color difference and the distance degree, respectively. Note that three RGB channels should be considered in the color similarity constraint c.

Fig. 6 shows the error map between the super-resolved result and the ground truth in the *StillLife* dataset. Fig. 6(c) and (b) are the close-up versions that use the hole-filling method or not. In these occlusion regions, the final result of the super-resolution is always inaccurate. We therefore estimate these values through their surroundings, and fortunately, the results are better. The punctiform area in 6(b) represents error estimation regions, and these regions decrease greatly through our hole-filling method in 6(c).

We obtain the difference maps $\mathcal{H}_i (i = 1, 2..., N)$ after propagating the central difference map $\mathcal{H}$ to all the side views and estimating the difference values in the hole regions. This difference map (Fig. 2(f)) indicates the missing high-frequency information between the $\mathcal{M}_i^l$ views and the desired high-resolution side views. Specifically, $\mathcal{H}_i$ is added to the VDSR super-resolved version $\mathcal{M}_i^l$ to form the final super-resolved results (Fig. 2(g)).

IV. Experimental Results and Discussions

In this section, we evaluate the proposed framework on several datasets including real-world scenes, microscope light field data and synthetic scenes. We compare our work with bicubic interpolation, the typical SISR method VDSR [10] and the PaSR method proposed by Boominathan *et al.* [11]. We also compare our work with the state-of-the-art iPADS method proposed by Wang *et al.* [12] using their datasets. The datasets we use for the evaluation contain several challenging scenes such as complex occlusion regions, scenes with large parallax and even challenging microscope light fields in dim lighting conditions. The quality of the reconstructed views is measured by the PSNR and structural similarity (SSIM) [35] against the ground truth images. SSIM produces a value between 0 and 1, where 1 indicates perfect perceptual quality with respect to the ground truth. In addition, we demonstrate how a reconstructed light field can be applied to enhance the depth estimation.

For light field datasets, we evaluate these methods according to two different scale factors: $\times 4$ and $\times 8$. We keep the central image of the light field unchanged and down sample the rest of the images as side view images. The central image is regarded as the the reference image while the original inputs of the side view images act as the ground truth for computing PSNR and SSIM.

We evaluate the bicubic interpolation, VDSR [10], PaSR [11] iPADS [12] and our approach on the different datasets. For SISR, we use the released model for super-resolution in scale $\times 4$, while the scale $\times 8$ VDSR results are obtained by applying $\times 4$ VDSR upsampling followed by $\times 2$ bicubic upsampling. For PaSR [11] in scales $\times 4$ and $\times 8$, we set the patch size in the L_i as 8×8, search range as 15 pixels and $\frac{1}{2\sigma^2} = 0.0125$, according to the paper.

A. Synthetic Scenes

We use the synthetic light field data from the HCI datasets [34] in which the angular resolution is the same as the original inputs (9×9). We use a cross-resolution input light field with two different super-resolution scaling factors ($\times 4$ and $\times 8$, respectively) to evaluate the performance of the proposed framework. The spatial resolution of the original light field images is the same as the ground truth.

Table I lists a quantitative evaluation on the synthetic dataset of the proposed approach compared with other methods, with scaling factors $\times 4$ and $\times 8$. Our approach achieves the best performance among all methods with the different scaling factors.

Fig. 7 shows the super-resolution results of different approaches by the scaling factor $\times 8$. As can be observed, our approach produces much sharper edges than other methods without any obvious artifacts across the image. In the *Buddha* case, the point of the dice is shown in the blue box. Our HCSR method is able to produce more accurate high-frequency detail information. The result of the PaSR method [11] has ghosting artifacts. The result of the VDSR method [10] has deformation; the point changed to an oval. The bicubic interpolation method generates a result with severely blurred artifacts. In the close-up version of our proposed approach, it is possible to observe increased sharpness and details in the super-resolved results. The results of Table I and Fig. 7 also indicate that the proposed scheme produces the least amount of artifacts.

TABLE I
QUANTITATIVE RESULTS (PSNR / SSIM) OF RECONSTRUCTED LIGHT FIELDS ON THE SYNTHETIC SCENES OF THE HCI DATASETS [34]

Dataset	Scale	Bicubic PSNR/SSIM	VDSR [10] PSNR/SSIM	PaSR [11] PSNR/SSIM	HCSR(Ours) PSNR/SSIM
Buddha	×4	31.3025 / 0.9976	32.8162 / 0.9846	32.0326 / 0.9815	**35.7420 / 0.9922**
	×8	27.4507 / 0.9435	28.5385 / 0.9575	27.7893 / 0.9505	**32.6646 / 0.9837**
MonasRoom	×4	32.0135 / 0.9829	34.1141 / 0.9896	38.5205 / 0.9963	**40.0340 / 0.9971**
	×8	28.3801 / 0.9593	29.7735 / 0.9712	34.2105 / 0.9900	**36.2754 / 0.9930**
StillLife	×4	23.5971 / 0.9115	23.9496 / 0.9200	25.4685 / 0.9448	**31.0211 / 0.9837**
	×8	22.4118 / 0.8787	22.8924 / 0.8960	23.5386 / 0.9135	**29.4210 / 0.9771**

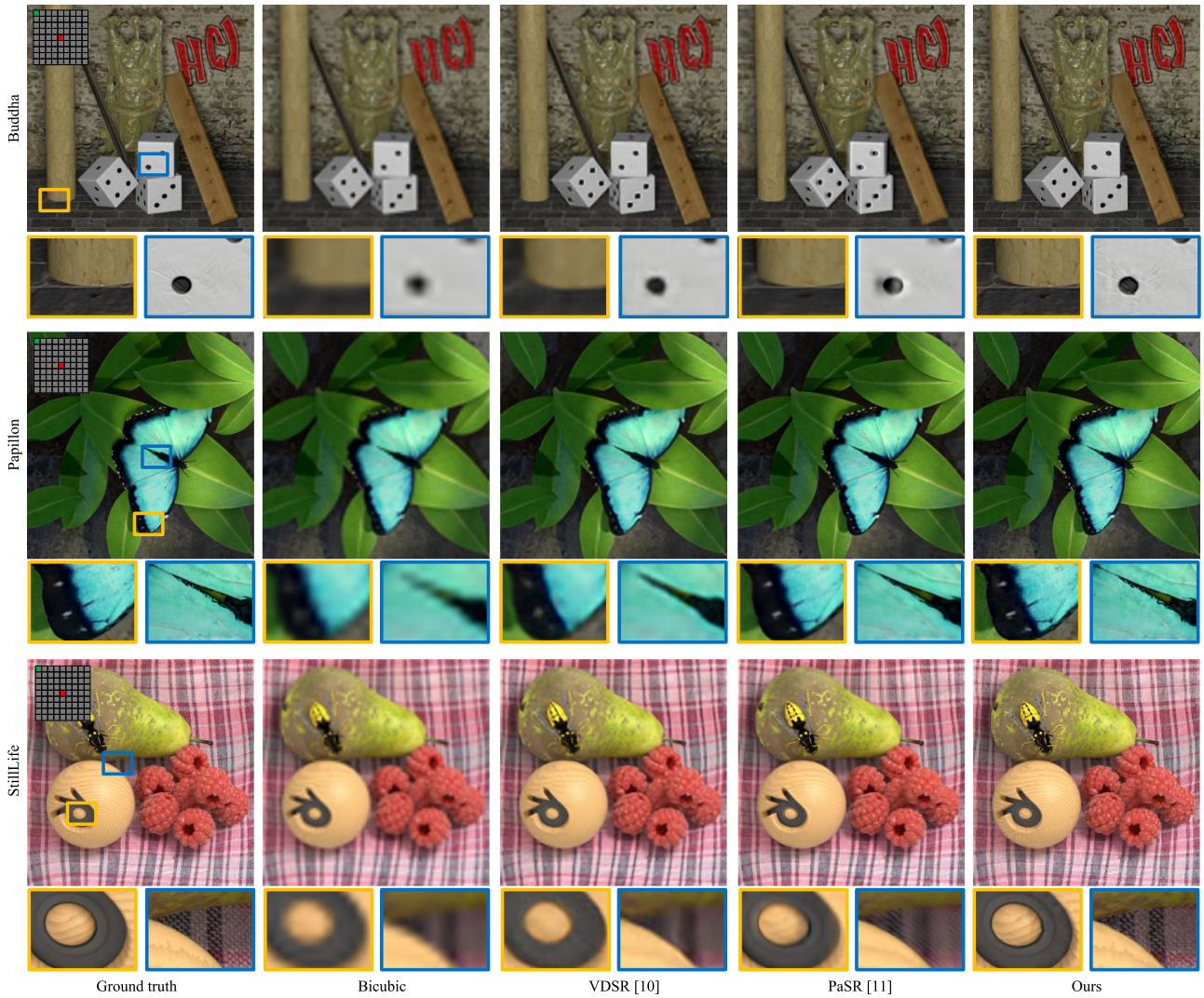


Fig. 7. Comparison of the proposed approach against other methods on the synthetic Scenes [34]. The results show the ground truth of reference images, the super-resolved images, and the close-up versions in the blue and yellow boxes.

B. *Real-World Scenes*

We evaluate the proposed approach using the scenes captured by a Lytro Illum camera from the Stanford Lytro Light Field Archive [6] and Stanford Light Field datasets [36]. For the Lytro light field, we reconstruct 7×7 light fields and select some representative scenes that contain occlusion and textureless regions. For the Stanford light field, we reconstruct a light field of 9×9 views.

1) Lytro Datasets: Table II lists the numerical results on the Lytro datasets. ×4 and ×8 represent that the reconstruct up-sampling factor is 4 and 8, respectively. The PSNR and SSIM values are averaged over the whole light field. Fig. 8 depicts some of the results (the up-sampling factor is 8) such as the *Plants* 12, *Leaves* and *Reflective* 29 scenes in the Stanford Lytro Light Field Archive. The patch-based approach by Boominathan *et al.* [11] is designed so that patches from each L_i image are matched with the center reference H_i image. Therefore, they

TABLE II
QUANTITATIVE RESULTS (PSNR / SSIM) OF RECONSTRUCTED LIGHT FIELDS ON THE REAL-WORLD SCENES CAPTURED BY LYTRO [6]

Dataset	Scale	Bicubic PSNR/SSIM	VDSR [10] PSNR/SSIM	PaSR [11] PSNR/SSIM	HCSR(Ours) PSNR/SSIM
*Plants*12	×4	28.2444 / 0.9200	31.1447 / 0.9606	31.7925 / 0.9673	**33.7690 / 0.9791**
	×8	26.9437 / 0.8847	27.1650 / 0.8882	27.3833 / 0.8891	**31.9264 / 0.9715**
Leaves	×4	25.8873 / 0.9758	30.0281 / 0.9908	31.1716 / 0.9920	**34.0150 / 0.9966**
	×8	23.5478 / 0.9572	23.8644 / 0.9583	23.9978 / 0.9586	**27.0761 / 0.9783**
*Reflective*29	×4	27.6793 / 0.9477	26.6961 / 0.9366	29.0102 / 0.9521	**36.5569 / 0.9924**
	×8	25.6079 / 0.9110	26.3488 / 0.9215	27.1989 / 0.9383	**34.2338 / 0.9885**

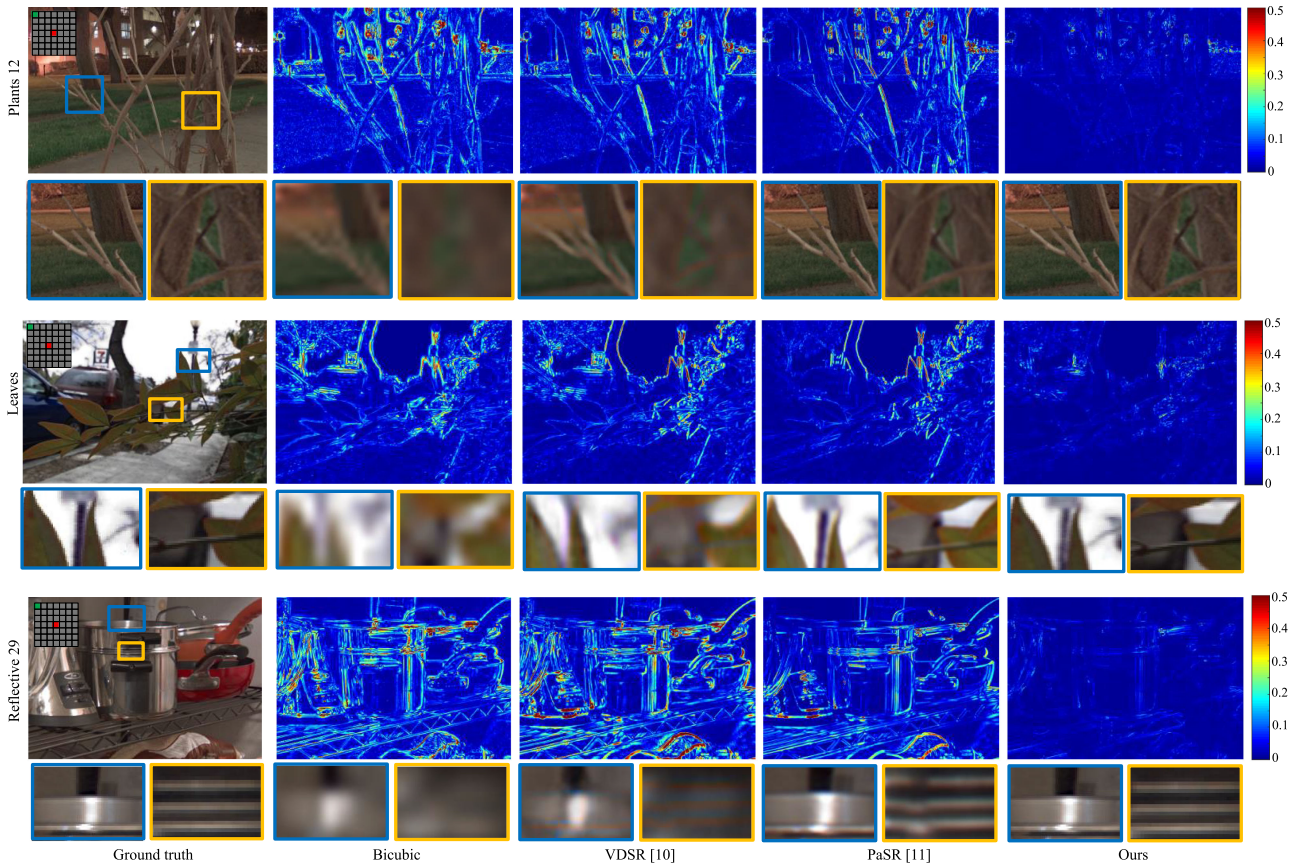


Fig. 8. Comparison of the proposed approach against other methods on the real-world scenes captured by a Lytro [6]. The results show the ground truth images and the error maps of the super-resolved results, and the close-up versions of the image portions in the blue and yellow boxes.

achieve a better performance than the other SISR method, bicubic interpolation and VDSR [10]. However, their approach cannot provide reasonable matching pairs in some specific regions, such as specular surfaces, and thus tends to fail in the textureless surface in the *Reflective* 29 case. Among these Lytro light field scenes, our proposed framework is significantly better than other approaches.

The *Plants* 12 case contains complicated occlusions and gloomy setting that make it challenging. The VDSR [10] and PaSR [11] results are quite blurry around the occluded regions such as the branches. The *Leaves* case includes some leaves with a complex structure in front of a street. The case is challenging due to the overexposure of the sky and the occlusion around the leaves, shown in the blue box. The result by bicubic interpolation shows serious blurring, the VDSR [10] result shows blurring artifacts around the leaves, and the PaSR [11] result contains ghosting artifacts. The *Reflective* 29 case is a challenging scene because of the textureless surfaces of the pot and the kettle. As demonstrated in the error maps and the close-up images of the results, the proposed approach achieves high performance in terms of super-resolved details and visual coherency.

2) Stanford Light Field Datasets: The Stanford light field dataset [36] contains light fields captured by a light field gantry system and thus has higher spatial resolution than light fields captured by Lytro. As seen in Table III, our approach produces results that are significantly better than other methods in both the scaling factors ×4 and ×8. We show four of these Stanford light field scenes in Fig. 9. The *Tarot Cards* scene contains a complex structure that makes it hard for the other approaches

TABLE III
QUANTITATIVE RESULTS (PSNR/SSIM) OF RECONSTRUCTED LIGHT FIELDS ON THE REAL-WORLD SCENES PROVIDED BY THE STANFORD LIGHT FIELD DATASET [36]

Dataset	Scale	Bicubic PSNR/SSIM	VDSR [10] PSNR/SSIM	PaSR [11] PSNR/SSIM	HCSR(Ours) PSNR/SSIM
TarotCards	×4	26.0773 / 0.9573	28.0312 / 0.9743	27.4733 / 0.9703	**33.3579 / 0.9924**
	×8	22.6408 / 0.8996	23.9477 / 0.9272	27.7344 / 0.9533	**30.2969 / 0.9835**
LegoKnights	×4	30.9126 / 0.9907	33.6056 / 0.9951	31.5034 / 0.9920	**36.0596 / 0.9974**
	×8	26.9367 / 0.9766	28.9750 / 0.9811	29.0602 / 0.9856	**31.5471 / 0.9914**
EucalyotusFlowers	×4	32.2410 / 0.9856	33.4495 / 0.9871	33.0404 / 0.9882	**35.0673 / 0.9919**
	×8	29.8656 / 0.9749	30.7956 / 0.9777	30.4894 / 0.9788	**33.1605 / 0.9872**
Amethyst	×4	32.6559 / 0.9904	34.3501 / 0.9924	34.8299 / 0.9942	**36.0025 / 0.9954**
	×8	29.0107 / 0.9777	30.0956 / 0.9823	31.6159 / 0.9878	**33.3072 / 0.9913**

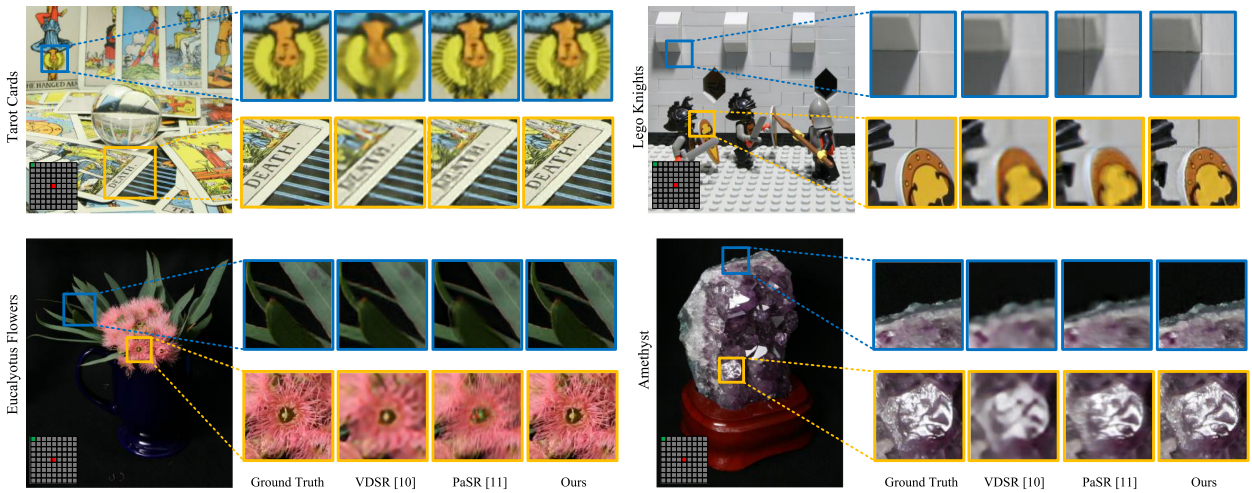


Fig. 9. Comparison of the proposed approach against other methods on the real-world scenes provided by the Stanford light field dataset [36]. The results show the ground truth of reference images and the super-resolved images, and the close-up versions in the blue and yellow boxes.

to accurately estimate the details in the images. However, our approach produces a plausible result that is reasonably close to the ground truth image. Note, for example, that only our approach is able to reconstruct the hair (blue box) clearly. The *Lego Knights* scene has complex structures with a very large parallax, where the maximum disparity of the adjacent viewing angle is 4 pixels. The PaSR [11] method is not able to handle the occlusion regions in which ghosting effects usually appear between the gap of the wall (blue box). The *Eucalyptus Flowers* scene contains a flower with complex occluded leaves. Our approach produces a reasonable result that is better than other methods. Despite the simplicity of the scene, the results of other methods are quite blurry in appearance, especially in the pistil (yellow box). Furthermore, their results contain tearing artifacts that can specifically be seen in the blue box. Note that only our approach is able to reconstruct all the details, such as in the pistil and in the occlusion boundaries. The *Amethyst* scene has some complex texture areas, such as in the yellow box. We produce a better result in these areas than the other methods relative to the ground truth.

C. *Microscope Light Field Dataset*

In this subsection, the microscope light field datasets captured by the camera-array-based light field microscope provided by Lin *et al.* [9] are tested. These datasets include challenging light fields such as complicated occlusion relations and translucency. The numerical results are tabulated in Table IV, and the super-resolved results are shown in Fig. 10. We reconstruct 5×5 light fields in the *Worm* case, *Electronic* case and the *Cotton* case respectively.

The *Cotton* case (Fig. 1) shows cotton fibers, which contain complicated occlusion regions. The result by VDSR [10] is quite blurry due to the large up-sampling factor. Although the result by PaSR [11] has a higher PSNR value, some ghosting artifacts still occur. The *Worm* case is more simply structured but contains transparent objects, such as the head of the worm. In these translucent regions, the PaSR [11] results are blurry because of the error match maps (yellow box). The *Electronic* case is in the dim lighting conditions. The regions marked by the two boxes contain blur regions and depth discontinuity regions. The results of all the other methods fail to reconstruct the detailed information. Among these challenging cases, our approach produces plausible results in these occluded, translucent regions and dim lighting conditions.

D. *Comparison With State-of-the-Art iPADS Method*

Based on the PaSR approach [11], Wang *et al.* [12] proposed a patch-based and depth-based synthesis method named iPADS.

TABLE IV
QUANTITATIVE RESULTS (PSNR/SSIM) OF RECONSTRUCTED LIGHT FIELDS ON THE MICROSCOPE LIGHT FIELD DATASETS [9]

Dataset	Scale	Bicubic PSNR/SSIM	VDSR [10] PSNR/SSIM	PaSR [11] PSNR/SSIM	HCSR(Ours) PSNR/SSIM
Worm	×4	38.6791 / 0.9993	39.2198 / 0.9994	39.3641 / 0.9994	**44.5872 / 0.9998**
	×8	33.0276 / 0.9973	34.6696 / 0.9982	35.3354 / 0.9984	**40.9682 / 0.9995**
Electronic	×4	39.9853 / 0.9954	40.9782 / 0.9958	47.1079 / 0.9994	**48.4650 / 0.9996**
	×8	38.6305 / 0.9931	39.8154 / 0.9949	37.4233 / 0.9908	**41.2975 / 0.9959**
Cotton	×4	37.3355 / 0.9954	38.0979 / 0.9966	38.0764 / 0.9966	**43.2985 / 0.9990**
	×8	31.3821 / 0.9837	34.6425 / 0.9923	32.5586 / 0.9876	**38.4137 / 0.9967**

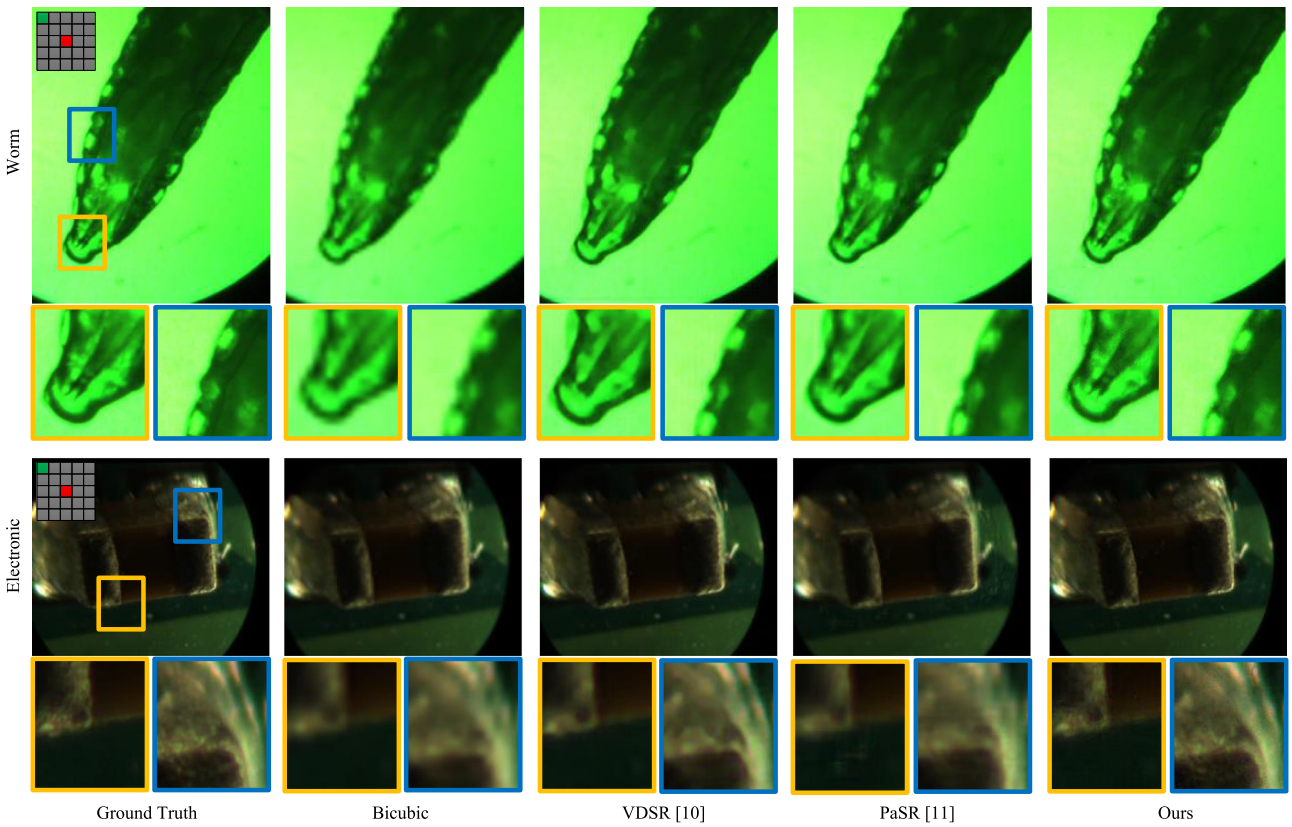


Fig. 10. Comparison of the proposed approach against other methods on the microscope light field datasets [9]. The results show the ground truth of reference images and the super-resolved images, and close-up versions in the blue and yellow boxes.

We use synthetic data sets for quantitative evaluation and real captured dataset from [12] for qualitative evaluation.

We use the Stanford and HCI datasets for synthetic evaluation. We chose 9 views from each light field, and the side-view images are also selected and downsampled in an 8-adjacency neighborhood of the central view. In terms of the number of images, the distance d between the central-view and the side-view is 3. In their experiment, the scaling factor of super-resolution is $\times 8$. We test our proposed approach based on their setup. The results of the comparison are listed in Table V. We can conclude that our approach achieves better performance in terms of both PSNR and SSIM.

We then use light field datasets which captured by the prototype [12] for qualitative evaluation. The data under real scenes is a central high-resolution image with 8 side views, and all the 9 views constitute the type of data that our proposed algorithm can directly apply. So in Fig. 11, we show some super-resolved results compared with iPADS method. From the results, our approach achieves better performance when compared with iPADS [12]. The proposed approach can recover better high-frequency details. For example, the yellow boxes in "Minions" and "Electrombile," the blue boxes in "Warning Sign" and

TABLE V
QUANTITATIVE RESULTS (PSNR/SSIM) OF OUR APPROACH AND THE IPADS METHOD [12]

Dataset	iPADS [12] PSNR/SSIM	HCSR(Ours) PSNR/SSIM
Couple	28.6465 / 0.9775	**29.9096 / 0.9908**
Maria	36.1043 / 0.9944	**37.5643 / 0.9960**
StillLife	26.4482 / 0.9602	**29.7796 / 0.9807**
TarotCards	31.6326 / 0.9873	**32.6982 / 0.9891**

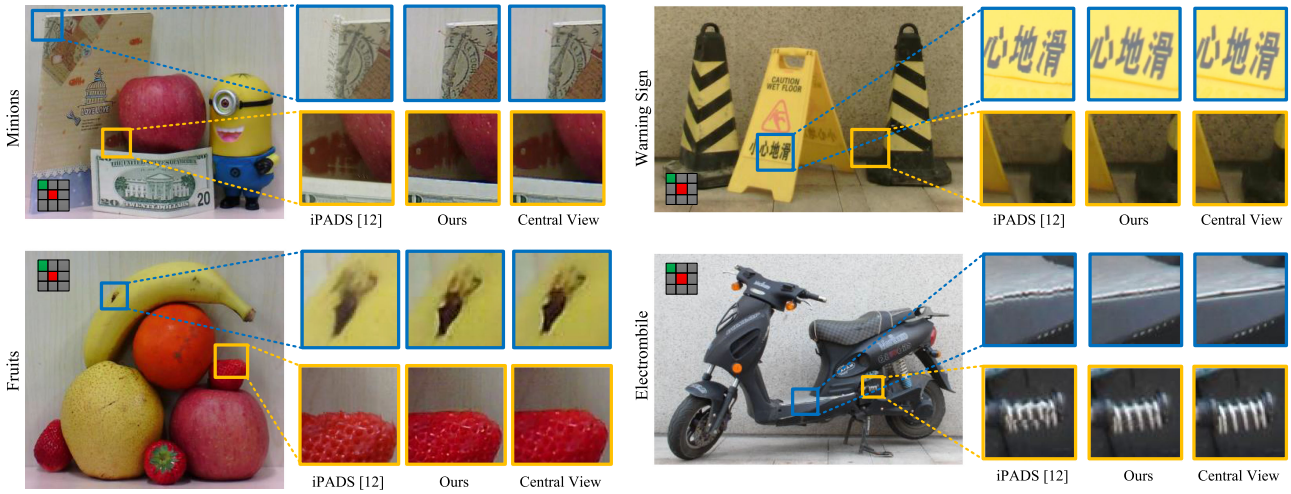


Fig. 11. The super-resolved results in real light field datasets using iPADS method [12] and our proposed method.

"Fruits". Moreover, in occlusion regions, the results of iPADS method suffer from tearing artifacts, while our scheme can handle the occlusion successfully, as further depicted by the blue boxes in "Minions," the yellow boxes in "Warning Sign" and "Fruits".

Finally, we test the computational cost compared with the iPADS method [12] under the same hardware condition. The algorithm was implemented in MATLAB 2016b. The computer is equipped with a GPU GTX 960 (Intel CPU E3-1231 running at 3.40 GHZ with 32 GB of memory). Super-resolving the *StillLife* case is taken as an example to test the computational efficiency. The iPADS method takes about 17 minutes to compute one super-resolved image of resolution 768×768 given an input image with a spatial resolution of 96×96. Compared to the iPADS method, our approach is more efficient in that it only takes 29 seconds for the same super-resolution setting. Both iPADS method and our VDSR module are implemented using Matlab without GPU operation. The running time for each module are as follows: Module 1 takes 9 seconds per view, and Module 2 takes 20 seconds per view.

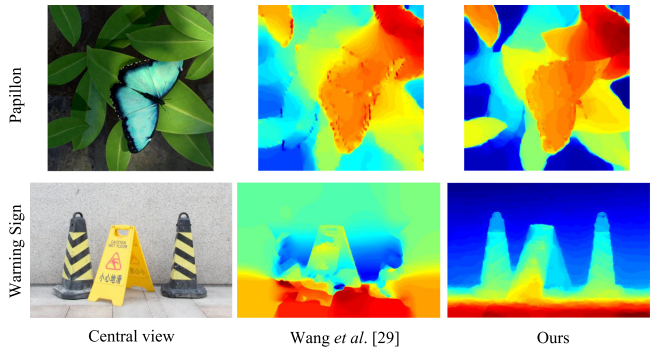


Fig. 12. The comparision of our propsed disparity estimation scheme agaist light field disparity estimation approach by Wang *et al.* [29].

TABLE VI
RMSE VALUES OF THE ESTIMATED DEPTH USING THE APPROACH BY WANG *et al.* [29] ON HCI DATASETS [34].

Dataset	Bicubic	VDSR [10]	PaSR [11]	Ours	Ground truth
Buddha	0.3272	0.2782	0.5497	0.1361	0.1098
MonasRoom	0.4511	0.3353	0.3572	0.2673	0.2631
StillLife	0.1651	0.1300	0.0737	0.0462	0.0408

E. Evaluation of Disparity Maps Blending

In this section, we compare our proposed disparity estimation scheme against a state-of-the-art light field disparity estimation approch by Wang *et al.* [29]. The input is cross-scale light fields with $\mathcal{SISR}$ images super-resolved by VDSR arranged around a HR central view. Our method uses only the cross-pattern images while Wang *et al.* [29] uses all the light field images. The results in Fig. 12 reveal that our disparity estimation scheme is more robust than Wang *et al.*'s approach on the cross-scale light fields, producing more clean and more accurate disparity map.

F. Application for Depth Enhancement

In this section, we demonstrate that the proposed light field super-resolution method can be applied to enhance depth estimation. The up-sampling factor is $\times 8$. We use the approach by Wang *et al.* [29] to estimate the depth of the scenes.

Table VI gives the RMSE values of the depth estimation results from using the method of bicubic interpolation, VDSR [10], PaSR [11], our approach and the ground truth light fields on the HCI datasets [34], including the *Buddha*, *MonasRoom*, and *StillLife* cases. Fig. 13 shows the depth estimation results on these scenes. The results show that our reconstructed light fields are able to produce more accurate depth maps that better preserve edge information than those produced by the other three methods; for example, the leaves in the flower pot in *MonasRoom* and the sheets behind the fruits in *StillLife*. Moreover, the enhanced depth maps are close to those produced by using the ground truth light fields.

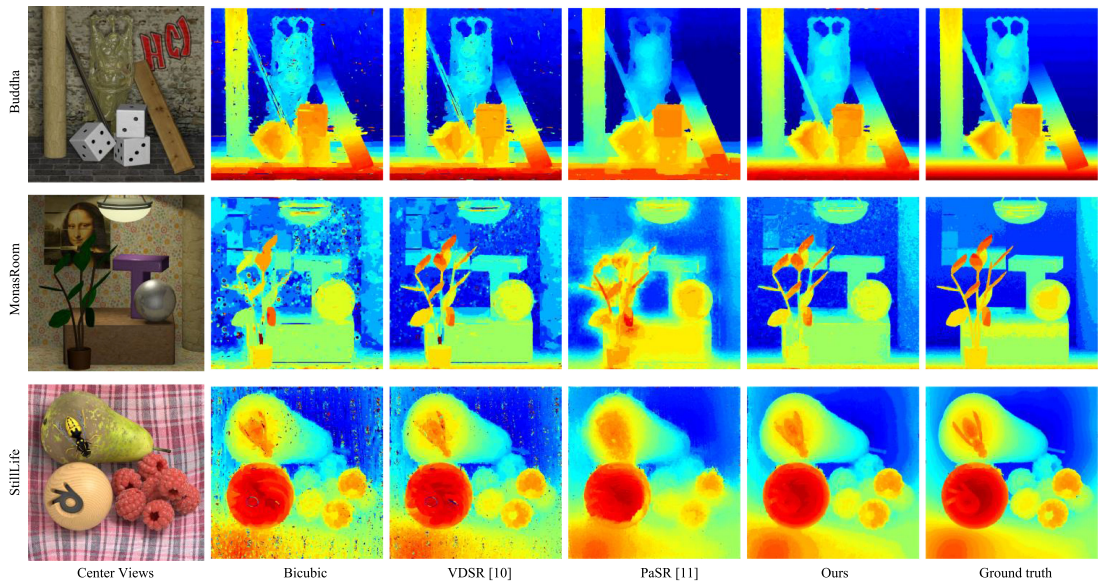


Fig. 13. The comparison of depth estimation using the reconstructed light fields according to the different methods.

V. Conclusion and Discussion

We have presented a light field super-resolution approach using a cross-resolution input, which consists of multiple side view images arranged around a central high-resolution image. By taking advantage of the light field structure, we have proposed the high-frequency compensation super-resolution (HCSR) scheme to compute the high-frequency information from the central image and warp this high-frequency information to all the side views. Such a step can achieve high-quality light field super-resolution using the cross-scale hybrid input. Experimental results on synthetic scenes, real-world scnes, and some challenging microscope light field datasets have demonstrated that our approach has excellent performance in the large scaling factor ($\times 8$) and parallax (up to 35 pixels) and has a higher computational efficiency compared to other methods.

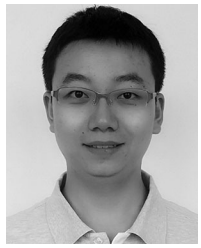
The main insight of our HCSR algorithm lies in that the target image in reference-based super-resolution (RefSR) can be decomposed into the "low frequency" components that can be obtained by VDSR, as well as the "high frequency" components that VDSR can hardly recover. The low frequency parts are obtained by super-resolving the input low-resolution image using VDSR, while the high frequency components are obtained by transferring those components from the high-resolution input image aided by the disparity information. With the proposed method, we refrain from the patch averaging operations in [11] and [12] that may blur the desired images, and successfully maximize the quality of the reconstructed image. It is worth mentioning that the modules in our scheme handle the super-resolution of each LR image independently, implying that our algorithm supports the data in [11] as well, if the calibration information of DSLR and the Lytro is provided. The only thing we need to modify is the disparity blending, to make it work for irregularly spaced cameras (by using the calibration result).

The main limitation of our proposed method may be that when non-Lambertian surfaces dominate in the scene, depth information cannot be calculated accurately and the overall super-resolution performance is compromised. The problem would be mitigated if reflectance analysis [37] were integrated using a data driven approach. In the future work, we can use data driven methods to learn the low frequency and high frequency priors of non-lambertian scenes, and integrate these learned prior into our pipeline for better reconstruction. Overall, we show that our method is robust and capable of being applied in general scenes even in the microscopic domain. We believe that such hybrid cross-scale input will be more popular in the future for its low data consumption because the data amount of the sum of the side view images is far less than that of the central views. Combined with the available angular super-resolution methods [20], [38], the hybrid cross-scale light field capture with the proposed HCSR approach would enable high-resolution free-viewpoint rendering and high-quality depth estimation under a very low bandwidth data consumption.

References

[1] A. Gershun, "The light field," *Studies Appl. Math.*, vol. 18, no. 1–4, pp. 51–151, 1936.
[2] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. 23rd Annu. Conf. Comput. Graph. Interactive Tech.*, 1996, pp. 31–42.
[3] G. Wu *et al.*, "Light field image processing: An overview," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 926–954, Oct. 2017.
[4] G. Lippmann, "Epreuves reversibles donnant la sensation du relief," *J. Phys. Theor. Appl.*, vol. 7, no. 1, pp. 821–825, 1908.
[5] B. Wilburn *et al.*, "High performance imaging using large camera arrays," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 765–776, 2005.
[6] "Stanford lytro light field archive," 2017. [Online]. Available: http://lightfields.stanford.edu/
[7] "Lytro," 2017. [Online]. Available: https://www.lytro.com/
[8] "RayTrix. 3D light field camera technology," 2018. [Online]. Available: http://www.raytrix.de/
[9] X. Lin, J. Wu, G. Zheng, and Q. Dai, "Camera array based light field microscopy," *Biomed. Opt. Express*, vol. 6, no. 9, pp. 3179–3189, 2015.
[10] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1646–1654.
[11] V. Boominathan, K. Mitra, and A. Veeraraghavan, "Improving resolution and depth-of-field of light field cameras using a hybrid imaging system," in *Proc. IEEE Int. Conf., Comput. Photography*, 2014, pp. 1–10.

[12] Y. Wang, Y. Liu, W. Heidrich, and Q. Dai, "The light field attachment: Turning a DSLR into a light field camera using a low budget camera ring," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 10, pp. 2357–2364, Oct. 2017.
[13] C. Y. Yang, C. Ma, and M. H. Yang, "Single-image super-resolution: A benchmark," in *Proc. Eur. Conf. Comput. Vis*, 2014, pp. 372–386.
[14] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 184–199.
[15] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 606–619, 2014.
[16] J. Li, M. Lu, and Z.-N. Li, "Continuous depth map reconstruction from light fields," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3257–3265, Nov. 2015.
[17] Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, and I. So Kweon, "Learning a deep convolutional network for light-field image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2015, pp. 24–32.
[18] Z. Zhang, Y. Liu, and Q. Dai, "Light field from micro-baseline image pair," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3800–3809.
[19] N. K. Kalantari, T.-C. Wang, and R. Ramamoorthi, "Learning-based view synthesis for light field cameras," *ACM Trans. Graph.*, vol. 35, no. 6, pp. 193–202, 2016.
[20] G. Wu, M. Zhao, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field reconstruction using deep convolutional network on EPI," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1638–1646.
[21] M. Ben-Ezra and S. K. Nayar, "Motion deblurring using hybrid imaging," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2003, pp. 657–664.
[22] X. Cao, X. Tong, Q. Dai, and S. Lin, "High resolution multispectral video capture with a hybrid camera system," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 297–304.
[23] E. Tola, Q. Cai, Z. Zhang, and C. Zhang, "Virtual view generation with a hybrid camera array," École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, Tech. Rep. 38, 2009.
[24] C. H. Lu, S. Muenzel, and J. Fleischer, "High-resolution light-field microscopy," *Comput. Opt. Sens. Imag.*, vol. CTh3B, pp. 23–27, 2013.
[25] G. Freedman and R. Fattal, "Image and video upscaling from local self-examples," *ACM Trans. Graph.*, vol. 30, no. 2, pp. 1–11, 2011.
[26] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 56–65, Mar./Apr. 2002.
[27] M. Muja, "Fast approximate nearest neighbors with automatic algorithm configuration," in *Proc. Int. Conf. Comput. Vis. Theory Appl.*, 2009, pp. 331–340.
[28] T. Wang, J. Zhu, N. K. Kalantari, A. A. Efros, and R. Ramamoorthi, "Light field video capture using a learning-based hybrid imaging system," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 133-1–133-13, 2017.
[29] T.-C. Wang, A. A. Efros, and R. Ramamoorthi, "Occlusion-aware depth estimation using light-field cameras," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 3487–3495.
[30] H. G. Jeon, J. Park, G. Choe, and J. Park, "Accurate depth map estimation from a lenslet light field camera," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1547–1555.
[31] L.C. , "Beyond pixels: Exploring new representations and applications for motion analysis," Ph.D. dissertation, Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 2009.
[32] Q. Yang, L. Wang, R. Yang, H. Stewenius, and D. Nister, "Stereo matching with color-weighted correlation, hierachical belief propagation and occlusion handling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 3, pp. 492–504, Mar. 2009.
[33] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1998, pp. 839–846.
[34] S. Wanner, S. Meister, and B. Goldlücke, "Datasets and benchmarks for densely sampled 4D light fields," in *Proc. Vision, Modeling Vis.*, 2013, pp. 225–226.
[35] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.
[36] "Stanford (New) Light Field Archive." [Online]. Available: http://lightfield.stanford.edu/lfs.html
[37] T.-C. Wang, M. Chandraker, A. Efros, and R. Ramamoorthi, "Svbrdf-invariant shape and reflectance estimation from light-field cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 5451–5459.
[38] N. K. Kalantari, T.-C. Wang, and R. Ramamoorthi, "Learning-based view synthesis for light field cameras," *ACM Trans. Graph.*, vol. 35, no. 6, pp. 193–202, 2016.



Mandan Zhao received the B.S. degrees from the School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou, China, in 2012 and the M.S. degrees from College of Aerospace Science and Engineering, National University of Defense Technology, Changsha, China, in 2014. He is currently working toward the Ph.D. degree at Zhengzhou Institute of Surveying and Mapping, Zhengzhou, China. His current research interests include light field analysis, image processing, and computer vision.

Gaochang Wu received the B.S. and M.S. degrees in mechanical engineering from Northeastern University, Shenyang, China, in 2013, and 2015, respectively. He is currently working toward the Ph.D. degree in control theory and control engineering at Northeastern University, Shenyang, China. His current research interests include data mining, signal analysis, image processing and computational photography.

Yipeng Li received the B.S. and M.S. degrees in electronic engineering from Harbin Institute of Technology, Harbin, China and the Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China, in 2003, 2005, and 2011, respectively. Since 2011, he has been a Research Associate with the Department of Automation, Tsinghua University. His research interests include computer vision, vision-based navigation UAS and three-dimensional reconstruction, and perception of general scene.

Xiangyang Hao received the Bachelor's, Master's, and Ph.D. degrees from Department of Photogrammetry and Remote Sensing, Zhengzhou Institute of Surveying and Mapping, Zhengzhou, China, in 1985, 1988, and 1996, respectively. He studied and was a Visiting Scholar with Institute of Photogrammetry and Cartography, Munich Bundeswehr University, Germany, from 2001 to 2002. He was elected as a member of the Chinese Society of Geodesy, Photogrammetry and Cartography in 1997. He is currently a Professor with Zhengzhou Institute of Surveying and Mapping. His research areas include computer vision, photogrammetry and navigation.

Lu Fang (SM'16) the B.E. degree from the University of Science and Technology of China, Hefei, China, in 2007 and the Ph.D. degree from Hong Kong University of Science and Technology, Hong Kong, in 2011. He is currently an Associate Professor with the Tsinghua-Berkeley Shenzhen Institute (TBSI), Shenzhen, China. His research interests include computational photography and visual computing. He was the recipient of NSFC Excellent Young Scholar Award, Best Student Paper Award in ICME 2017, Finalist of Worlds First 10 K Best Paper Award in ICME 2017, Finalist of Best Paper Award in ICME 2011, etc.

Yebin Liu (M'12) received the B.E. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2002, and the Ph.D. degree from the Automation Department, Tsinghua University, Beijing, China, in 2009. He was a Research Fellow with the Computer Graphics Group, Max Planck Institute for Informatics, Saarbrücken, Germany, in 2010. He is currently an Associate Professor with Tsinghua University. His research areas include computer vision and computer graphics.