

The Light Field Attachment: Turning a DSLR into a Light Field Camera Using a Low Budget Camera Ring

Yuwang Wang, Yebin Liu, Wolfgang Heidrich, and Qionghai Dai

Abstract—We propose a concept for a lens attachment that turns a standard DSLR camera and lens into a light field camera. The attachment consists of 8 low-resolution, low-quality side cameras arranged around the central high-quality SLR lens. Unlike most existing light field camera architectures, this design provides a high-quality 2D image mode, while simultaneously enabling a new high-quality light field mode with a large camera baseline but little added weight, cost, or bulk compared with the base DSLR camera. From an algorithmic point of view, the high-quality light field mode is made possible by a new light field super-resolution method that first improves the spatial resolution and image quality of the side cameras and then interpolates additional views as needed. At the heart of this process is a super-resolution method that we call iterative Patch- And Depth-based Synthesis (iPADS), which combines patch-based and depth-based synthesis in a novel fashion. Experimental results obtained for both real captured data and synthetic data confirm that our method achieves substantial improvements in super-resolution for side-view images as well as the high-quality and view-coherent rendering of dense and high-resolution light fields.

Index Terms—Light Field, Super-resolution, Computational Imaging.

1 INTRODUCTION

Recent years have witnessed a strong renewal of interest in capturing and reconstructing light fields (see, e.g., [1], [2]). The main driving force of this body of work is the realization that many interesting applications of light fields (e.g., [3], [4], [5]) are currently hindered by the low resolution and poor image quality of commercially available light field cameras. State-of-the-art light field camera designs attempt to take advantage of the redundancy in the light field data structure to achieve a better trade-off between the spatial and angular dimensions, by means of both hardware and software innovations. Some of these designs involve the modification of existing cameras and adding programmable apertures [6], [7] or coding modules [2] into traditional camera designs, and these approaches use compressive sensing to recover high-resolution spatial information. However, this kind of internal camera modification is difficult to build and inevitably compromises the resolution of the 2D imaging mode as well as other optical properties, such as the light efficiency in the case of coded modulation [2].

Meanwhile, discrete low-resolution cameras are continuing to decrease in both size and cost, which prompts us to propose a ring of many such cameras (8 in our prototype) that can act as a lens attachment for a central high-quality DSLR camera, thereby converting the DSLR camera into a light field camera without compromising its 2D imaging performance. We take advantage of



Fig. 1. Left: Conceptual design of our light field lens attachment. Right: The top left image is a scene captured using our system, and the top right image is the estimated disparity map. The bottom four images are: (a) an image patch from a side-view camera, (b) an image patch from the center-view DSLR camera, (c) the patch-based super-resolution result, and (d) the super resolution result obtained using our method.

the high spatial resolution of the DSLR image and the angular information provided by the side cameras, to devise a new light field super-resolution method that allows us to obtain a dense light field in which each view approaches the quality of the main DSLR camera view.

Specifically, this paper proposes a method called iterative Patch- And Depth-based Synthesis (iPADS) for light field super-resolution and synthesis using the low-resolution side views in combination with a central high-resolution DSLR image. The key idea of iPADS is to provide high-resolution patch candidates that are more similar to the high-resolution ground truth of the side-view images compared with the available central image. By utilizing depth information using a phase-based rendering approach, these patches are rendered from the central image to preserve the high-frequency detail of the central image.

Our proposed iPADS method therefore takes advantage of the natural complementarity between patch-based synthesis (ensuring low-frequency accuracy) and depth-based synthesis (preserving high-frequency detail) in an iterative optimization framework.

- Y. Wang, Y. Liu and Q. Dai are with Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: wang-yw14@mails.tsinghua.edu.cn; {liuyebin,qhdai}@tsinghua.edu.cn)
- W. Heidrich is with Visual Computing Center, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia (e-mail: wolfgang.heidrich@kaust.edu.sa).
- Corresponding Author: Yebin Liu (e-mail: liuyebin@tsinghua.edu.cn).
- This work was supported by the National key foundation for exploring scientific instrument No.2013YQ140517, the National NSF of China grant No.61522111 and No.61531014.

Experimental results demonstrate that the proposed iPADS method substantially improves the quality of the reconstructed high-resolution light field. In summary, our key technical contributions are as follows:

- We introduce a concept for a light field lens attachment that can convert a DSLR camera into a light field camera with high spatial and angular resolution, and we develop the first prototype to confirm its feasibility.
- We present an optimization framework called iterative Patch- And Depth-based Synthesis (iPADS) to achieve light field super-resolution using the data captured by our light field attachment. The proposed method iterates between patch-based synthesis for super-resolution and depth based synthesis for providing better patch candidates to achieve light field reconstruction with high spatial and angular resolution.
- We propose a new depth-based synthesis method that can jointly synthesize high-resolution side views using the texture of the central high-quality view and estimate the depth of the central view with high quality.

Our proposed light field camera bears some similarities to several other recent works. Similar to Boominathan et al [8], we synthesize a high-resolution light field from a high-quality 2D camera and a low-quality light fields. However, the light field camera considered in [8] is an integrated device (a Lytro camera), which increases the cost and limits the available baseline because all views share a main lens. By comparison, the side cameras in our method have independent optical systems, which can be inexpensive because of the low required resolution, and the whole system allows for much larger baselines. We also show that our super-resolution method is a significant improvement over that of [8]. Perhaps most similar to our concept is the upcoming commercial camera by Light¹, which consists of an array of discrete cameras mounted on a common body. Unfortunately, very few technical details are currently available about either the hardware or the software of this system. However, we note that from a hardware point of view, our design allows the use of high-quality lenses and very high-resolution sensors in the central DSLR, and that our super-resolution method is a clear improvement over the state-of-the-art. We believe that the proposed prototype and system can inspire future light field camera designs in pursuit of better and much more convenient light field imaging. The data and the code of this work will be made public.

2 RELATED WORK

Light fields [9], [10] provide a new angular dimension that enables various visual applications, such as depth estimation [1], [11], [12], scene refocusing [3], 3D display [4] and microscopy [13], [14]. Many recent works have attempted to capture or synthesize high-quality light fields from different types of input data.

Light field capture. In recent years, many light field capture methods have been proposed. These can first be divided into camera-array methods and single-camera methods. Camera-array methods [15], [16], [17] are capable of capturing light fields with both a high spatial resolution and a large angular range. However, the hardware required for these methods is expensive and complex, is nearly impossible to adapt for portable applications,

and requires considerable bandwidth, storage and power. Single-camera methods capture the angular information of a scene by sacrificing spatial resolution or using multiple shots. One subset of such methods involves adding a lenslet array [18], [19], or a mask [20] close to the sensor in a traditional camera architecture. A second subset of such methods is programmable aperture imaging [6]. Compressive sensing and dictionary learning have also been applied to reduce the number of images required by exploiting the self-similarity among light field images [2], [21]. The Pelican imaging camera array [22] is a portable, practical light field capture system that can be implemented in a cellphone but has a limited angular range. The concurrent work of Wang et al. [23] estimates depth by combining a stereo camera and a DSLR camera. In comparison to [23], our setup uses more side-view cameras of much lower resolution, with the aim at super-resolution of a sparse regular light field for dense light field interpolation.

Light field super-resolution. For single-camera light field imaging, the key disadvantage is its low spatial resolution. Several methods of restoring high-frequency information have recently been proposed. On the hardware side, the Lytro ILLUM [19] changes the original Lytro [18] setup by placing the lenslet array at different positions, and the Raytrix light field camera [24] employs a lenslet array with different focal lengths to achieve a higher spatial resolution. On the algorithm side, Bishop et al. [25] have proposed a method of estimating both a high-resolution depth map and light fields in the Bayesian framework using a prior based on a Lambertian texture. Other super-resolution methods using Bayesian inference [26] or variational optimization [27] can also effectively improve the spatial resolution of the light field. The resolution improvement is, however, typically below a factor of $4\times$. To achieve a significant improvement, an additional high-resolution reference image of the same scene is added as input; to this end, a hybrid imaging system [8] has been proposed based on a patch based synthesis strategy for synthesizing high-spatial-resolution light field images with a scale factor of $8\times$. However, the quality of the recovered light field images is not as good as that of the input high-resolution image. The high frequency spatial details are lost in the recovered super-resolution images.

Light field synthesis. The goal of light field synthesis is to improve the angular resolution and range of light fields. The available synthesis methods can be divided into two categories: depth-based methods and phase-based methods. The former mainly depend on an explicit geometry [28] and are sensitive to errors in the disparity estimation. The latter methods rely less on depth but requires a denser sampling of the light field as input [29] [26].

Signal processing frameworks using phase-based analysis [30] or Fourier analysis [31] have also been proposed recently for the synthesis of light field images using only a few or a small subset of light field images. Depth-aided phase-based synthesis (DAPS) [32] combines the advantages of both depth-based and phase-based method by integrating the depth information into the phase-based processing framework to compensate for the spatial effects of floating texture deformation across views. Although the visual quality of such phase-based synthesis methods is high because they preserve the high-frequency image components, the recovered images are not guaranteed to be well aligned with the ground truth if errors arise during depth recovery. In this paper, we use DAPS in our depth-based synthesis module.

1. <http://www.light.co/camera>

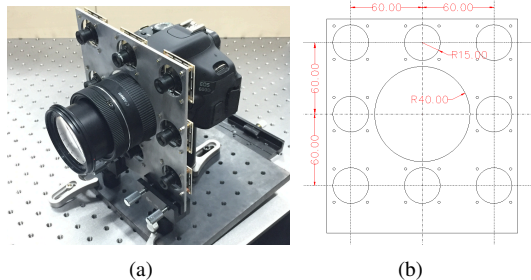


Fig. 2. Light field attachment prototype: (a) prototype; (b) design of the camera layout. Units: millimeters.

3 PROTOTYPE CAMERA SETUP

We place 8 off-the-shelf low-end USB cameras on a custom aluminum mount around a high-resolution camera, enabling the sparse light field capture of 3×3 views, as shown in Fig.2. All side-view USB cameras are placed in a regularly spaced square grid with a distance of 60mm between neighboring cameras. For the central camera, we use a Canon 600D DSLR camera with a Canon EF-S18-135 mm $f/3.5-5.6$ IS DSLR lens. We crop the input images to the region of overlap among all cameras before further processing. For scenes at far distance, this overlap region will approach the full sensor resolution, but for the more interesting case of close scenes considered in this paper, the central image has a resolution of 2944×1808 after boundary cropping, whereas the resolution of the side-view images is 368×226 , which is exactly $1/8$ of the resolution of the central DSLR view in both the horizontal and vertical dimensions. The goals of this work are to improve the resolution of all the side views and synthesize a dense light field with the same resolution as that of the central view.

We fix the focus settings for all cameras and calibrate the intrinsic parameters of the entire system using a standard camera calibration toolbox [33]. Then, bundle adjustment software [34] is applied to calibrate the extrinsic parameters of all cameras with reference to the central camera. Finally, we rectify each side-view image by projecting it onto a reference plane parallel to the DSLR image plane using homography warping, such that all cameras share the same rotation. Note that although the cameras are not ideally regularly displaced, the axial displacement errors are smaller than 3mm and the horizontal and vertical misalignments are smaller than 1mm ; thus, any errors in the image space are on the sub-pixel scale, and our algorithm is robust to these errors.

4 ALGORITHM OVERVIEW

As shown in Fig. 3, given the captured data from Sect. 3, the proposed algorithm for the synthesis of high spatial and angular resolution light fields can be divided into two main steps: iPADS-based super-resolution of the side-view images with optimized depth estimation and the synthesis of a dense light field. The iPADS algorithm iterates between the patch-based super-resolution on the side-view images (step 1 in Fig. 3), depth-based image synthesis (steps 2,3 and 4 in Fig. 3) and updating the patch database (step 5 in Fig. 3).

Specifically, iPADS is implemented by iterating the following steps: 1) Build a feature dictionary for image patch matching using the given high-spatial-resolution image, then perform patch-based super-resolution (PaSR) [8] on the low-resolution side-view images using this patch-based feature dictionary. In this step, some of the high-frequency texture components are lost because of

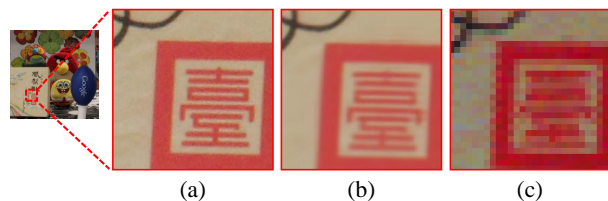


Fig. 4. Loss of high-frequency information in PaSR: (a) ground truth; (b) super-resolution result of PaSR; (c) low-resolution input.

variations in scene depth. 2) Perform multi-view depth estimation based on the central-view image and the 8 improved-resolution side-view images. 3) Apply an improved version of the depth-aided phase-based synthesis (DAPS) approach [32] to render new side-view images and perform a depth estimation update. During this step, most of the high-frequency texture information is restored. 4) Apply an optical-flow-based warping algorithm to rectify synthesis position errors that may have arisen in step 3 as a result of depth estimation errors. 5) Add the warped views into the patch-based feature dictionary. During this step, high-spatial-resolution images from other views are added to the dictionary to act as candidates for the next iteration of PaSR.

5 IPADS SUPER-RESOLUTION

This section introduces the proposed iterative Patch- and Depth-based Synthesis method for achieving super-resolution of the side-view images and performing depth estimation of the central-view image. We first introduce the two main components of the process followed by a description of the iteration scheme.

5.1 Patch-based super-resolution

We adopt a previously developed patch-matching-based super-resolution [8] method as the first step of our iPADS algorithm. Patch-matching-based super-resolution relies on the similarity between the input high-resolution image and low-spatial-resolution light field. The method first builds a dictionary from the given high-spatial-resolution image patches and then uses first- and second-order derivative filters to extract the feature of each high-spatial-resolution patch [35].

We use the same parameter settings applied in [8]. The patch sizes of the low- and high-resolution patches are 8×8 and 64×64 respectively. During the first iteration, we use the same dictionary for each side view, which is constructed from the center-view DSLR image. During subsequent iterations, we build different dictionaries for different side-view images using the center-view DSLR image and the corresponding synthesized super-resolution side-view images. These synthesized side-view images feature a similar visual quality as the central input image, but with improved parallax information corresponding to the desired side views.

However, this method causes high-frequency information to be lost in the super-resolution images. Note that the super-resolution versions of the low-resolution patches are generated by densely sampling the low-resolution image with only one-pixel shifts to ensure the existence of overlap areas among the super-resolution patches. For these areas, the average value of all overlapping super-resolution patches is computed as the final pixel value. As shown in Fig. 4(b), high-frequency details are lost because of this average filtering.

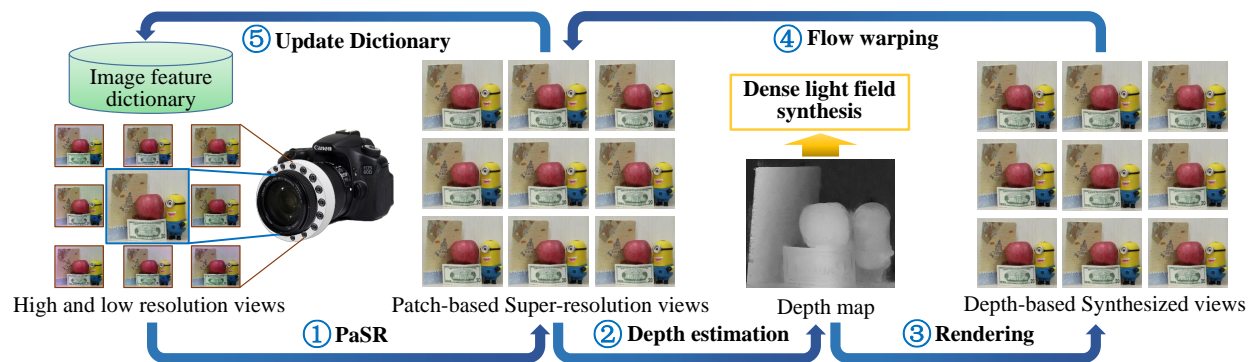


Fig. 3. Overview of the pipeline of the proposed iPADS algorithm and dense light field synthesis. iPADS iterates between patch-based super-resolution (PaSR, step 1) and depth-based synthesis (step 2, 3 and 4) for the synthesis of new images that are then updated into the patch database (step 5) for better patch-based super-resolution. The depth-based synthesis includes multiview depth estimation (step 2), depth-aided phase-based synthesis (DAPS, step 3) and optical-flow-based warping (step 4). Finally, a dense light field synthesis step is executed to render the final output using the estimated depth information.

5.2 Depth-based synthesis

Using the high-resolution central-view image and the side views after patch-based super-resolution, we perform depth-based synthesis to render high-resolution side views from the central image. Unlike patch-based synthesis, depth-based synthesis [36] using a single reference image usually preserves high-frequency details, although the performance of such methods actually depends on the accuracy of the depth information. To mitigate synthesis errors and artifacts due to inaccurate depth estimation, we apply three sequential processing steps in this module to ensure high-quality view synthesis and depth estimation. We first estimate a disparity map following a multi-view disparity estimation pipeline, which provides a rough but robust initial depth estimation. Then this initial depth map is refined using the recently proposed depth-aided phase-based synthesis (DAPS) method [32], which simultaneously enables synthesis with high-frequency details for the target side views. Finally, optical-flow-based warping is applied to compensate for residual mismatching. Through these three steps, we successfully obtain side-view images that are as accurate as possible in which high spatial frequency detail is preserved. The results exploit the angular information provided by the side views to obtain better candidate patches compared with the center view for patch-based super-resolution.

Multi-view disparity estimation Using all 3×3 views at the same resolution, we modify the existing cost volume filtering approach for binocular disparity estimation [37] by summing all of the cost-volumes of the 8 center-side stereo pairs. Based on the observation that the side-view images are obtained from patch-matching super resolution with missing high frequency details, we adopt a coarse-to-fine stereo matching strategy by first down-sampling the 9 view images by a factor of $1/4$ to estimate the initial disparity map and then iteratively up-sampling the disparity map by a factor of 2 for further disparity map refinement. For the cost-volume at each pixel position, we search only for the local minimal cost near the disparity value obtained at the previous level of granularity.

Depth-aided phase-based synthesis After calculating the initial disparity map for the central view, we apply the recently proposed DAPS [32] method to jointly synthesize high-quality side-view images and iteratively improve the disparity map. The central view, the synthesized side views obtained in PaSR step, and the disparity map estimated in the previous step are fed into DAPS to improve the quality of the side views. DAPS includes a phase-

based image synthesis strategy that can warp from an input image to a nearby view based on the disparity information. DAPS outperforms traditional depth-image-based rendering (DIBR) methods in two respects. First, DAPS takes advantage of both complex steerable pyramid filters [38] for signal reconstruction and a depth-based prior to guide the reconstruction of the spatially varying effects of floating texture deformation; thus it can achieve coherent results in the rendering of different views (see the comparison with a typical DIBR method [39] presented in the supplemental video). Second, an iterative disparity refinement scheme is incorporated into the DAPS framework to improve the quality of both the disparity calculations and the synthesized views.

The original DAPS simply synthesizes novel views (the side views in our case) using only the texture from source image, i.e., the central view. In our case, since the baselines of the central image and the side images are much wider, simply filling in the disoccluded regions would result in obvious artifacts. To mitigate these artifacts, we leverage the available PaSR-synthesized side view image and trade off between the DAPS-synthesized image. We first compute a gradient map of the disparity map and then enforce a Gaussian filter on this gradient image. This filtered gradient map encodes the weights that are used in the linear combination of the DAPS-synthesized images and PaSR-synthesized side-view images. Actually, larger filtered gradient implies that the texture of the boundary pixels should originate from the PaSR result. Fig.10 illustrates the immediate result of the filter gradient map (b) and the blended result (c), the latter of which shows substantial improvement over the DAPS-synthesized result (a).

To make full use of the multi-view side-view images, we apply DAPS to each central-side view pair one by one. After the DAPS processing for one pair is completed, the disparity map for the central view is refined and is used for the processing of the next image pair. Fig.5(d) illustrates how the quality of the disparity map is improved after DAPS refinement. Compared with the disparity map before DAPS (Fig.5(c)), the PSNR increases by 2.4 dB.

Optical-flow-based warping After DAPS, the high-frequency details are perfectly recovered in the synthesized side-view images. However, because of inevitable errors in the disparity map, the synthesized images are slightly inconsistent with the ground truth. To remove these local mismatches, we therefore use the optical flow methods described in section 2.3.6 of [40] to warp the DAPS-synthesized images to the patch-based super-resolution images. Since the misalignments are small, the optical-flow-based approach can effectively remove them. To evaluate this

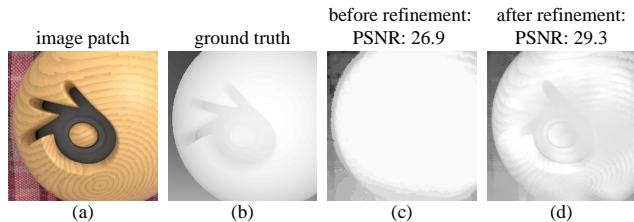


Fig. 5. Results of DAPS refinement: (a) image patch; (b) a ground-truth disparity map; (c) disparity map before refinement; (d) disparity map after refinement.

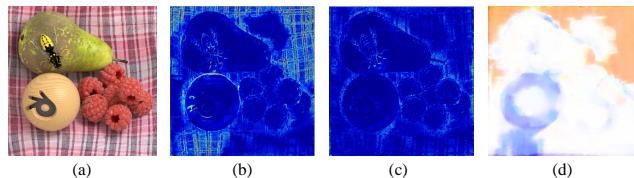


Fig. 6. Results of optical-flow-based warping refinement: (a) ground truth; (b) error map before warping; (c) error map after warping; (d) color-coded optical flow used for warping.

flow-based warping step, we present the error maps before and after warping for a synthetic data set. We use the default parameter settings in [40]. As Fig. 6 shows, the error after flow-based warping is much lower. The DAPS synthesized-images contain high-frequency details recovered from center-view high-resolution images, and the patch-based super-resolution images contain more angular information. Through this optical-flow-based warping refinement, we are able to reduce the local inconsistencies between the synthesized images and the ground truth. Note that it is not reasonable to directly warp the central-view image to the patch-matching-based super-resolution images. This is because the disparities between the central view and the side views are larger and the optical flow method is not robust to different kinds of textures and can easily become trapped in a local optimum.

Fig.7 shows how an image patch evolves during a single iteration. Compared with the ground truth, the patch after PaSR lost high frequency details. After DAPS synthesis, the high-frequency textures are restored, and the PSNR increases. However, there are still mismatches with the ground truth because of errors in the inferred disparity map. The warping step achieves another PSNR improvement, even though the high resolution details are a little bit lost.

5.3 Iterative optimization

After optical-flow-based warping, we add the synthesized high-resolution side-view images into the patch-based dictionaries. Compared with the central-view image, these synthesized side-view images share a similar image quality but also represent the parallax differences arising from the offset camera positions. Therefore, the synthesized side-view images are a better source of candidate patches for the PaSR step in the next iteration. Here, we build different dictionaries for different side-view images using the corresponding synthesized images.

We then iterate the entire iPADS algorithm 4 times to achieve convergence of synthesized side views and the estimated depths. The improved dictionary leads to a better patch-based super-resolution quality in terms of high-frequency recovery, and this improved preservation of the high-frequency components in turn leads to better depth estimations and depth-based synthesis. Fig. 8 shows an example of the improvement in the disparity maps generated with an increasing number of iterations. We also

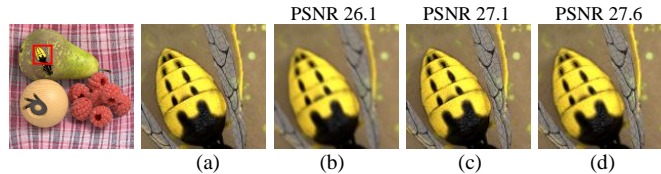


Fig. 7. Evolution of a patch during a single iteration: (a) ground truth; (b) after PaSR; (c) after DAPS; (d) after optical-flow-based warping.

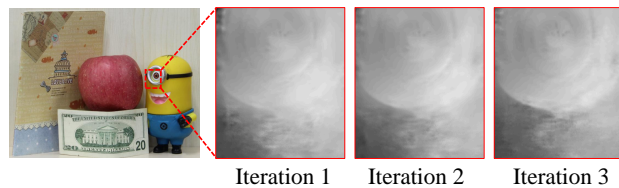


Fig. 8. Iterative refinement of the disparity map.

quantitatively evaluate our method based on the PSNR metric on two data sets, maria and still life, for which ground truth depth information is available. As Fig. 9 shows, the PSNR value increases with an increasing number of iteration on both data sets. The PSNR of the depth map is computed in terms of the disparity (in pixels), and linearly normalize the ranges of the ground-truth disparity map and our computed disparity map to $[0,255]$, and both of these two normalization operations use the same minimum and maximum values. Thus the peak value used to compute the PSNR is 255, which is the same as that for RGB images.

6 DENSE LIGHT FIELD SYNTHESIS

To synthesize a dense light field, the DAPS method can be adopted to synthesize images from arbitrary viewpoint close to the central viewpoint using only the central view-image and its disparity map. However, as stated in Sect. 5, disoccluded areas in the target view are not well handled in the original DAPS framework. In our case, the recovered side views provide angular information that can be used to fill in these disoccluded areas. We assume that the disparity of a disoccluded pixel is the same as the disparity of the closest background pixel on the boundary, then this pixel in the disoccluded area can be traced and copied from the side-view images. Using this filling strategy, we can achieve high-quality dense light field synthesis. Fig. 10 shows the quality improvement for a disoccluded area in an in-between target view, and our result with filling dis-occluded area looks much better.

7 RESULTS AND DISCUSSION

We evaluate the performance of iPADS for side views and dense light field rendering on the Stanford [41] light field data set, the HCI [42] light field data set, and indoor and outdoor data captured using our hardware system.

Experimental design For the Stanford and HCI data set, we select 9 views from each light field with a layout similar to that of our data capture system. Because these light fields have a much higher angular resolution compared with ours, we select the side-view images with $d = 3$ (distance from the central view in terms of the number of images) in an 8-adjacency neighborhood of the central view. The input low-resolution side-view images are obtained by down-sampling each image by a factor of $1/8$, and the original high-resolution images serve as the ground truth. For patch-based super-resolution processing, we use the same settings applied in [8].

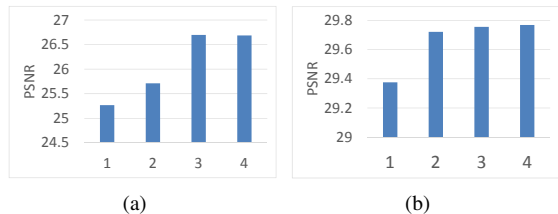


Fig. 9. PSNR values for the depth maps generated in different iterations: (a) maria; (b) still life

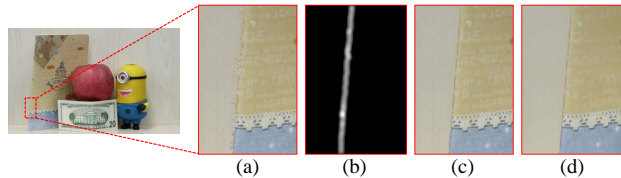


Fig. 10. Handling disoccluded area: (a) DAPS-synthesized view; (b) filtered gradient map; (c) result of filling in the disoccluded area using the side view image; (d) result of filling in the disoccluded area on one of the dense interpolated image.

Super-resolution results We first evaluate our method on four light fields: maria, still life, couple, and tarot. The PSNR values of the PaSR images and our depth-based synthesis images in each iteration are shown in Fig. 11. Each PSNR value is obtained by averaging over the PSNRs of all side views. The PSNR values obtained using both our method and PaSR increases with an increasing number of iterations, and the PSNRs for our method are higher than those for PaSR in each iteration. After all iterations are completed, our method yields an improvement of approximately 3dB compared with the PaSR values for the results obtained without iteration (i.e., in iteration 1).

In each iteration, the average PSNR value of our results is higher than that of the PaSR results because our depth-based synthesis method can more effectively recover high-resolution texture information from the center-view image. Because of the dictionary updates, better candidate patches are added into the dictionaries, contributing to the enhancement of the PaSR results and allowing more angular information from the low-resolution images to be exploited. With more angular information, the results of our depth-based synthesis also improve. We observe that the improvement in the PSNR value is small after 4 iterations, and thus fix the number of iterations to 4 for all further experiments.

We also evaluate the super-resolution results based on another metric, the SSIM index [43], which emphasizes the structural similarity between images. As Fig. 10 shows, our method results in a larger improvement on all four data sets. For each data set, the SSIM value increases with an increasing number of iteration. Moreover, this increase is more significantly on data sets with large depth variations, such as the still life data set. This is because for a data set with large depth variations, PaSR will introduce blur and structure error, whereas our method attempts to restore structural information.

Fig. 12 shows several super-resolution patches cropped from the four simulations. It is obvious that the patches generated by our method contain better high-frequency details than those generated by PaSR, especially for patches with large depth variations. Fig. 13 shows super-resolution patches cropped from two of our captured data sets. Again, our results are much better compared with those of PaSR. Moreover, a comparison with the high-resolution center-view image reveals that our method preserves almost all of the high-resolution details. More super-resolution results are given in the supplemental materials.

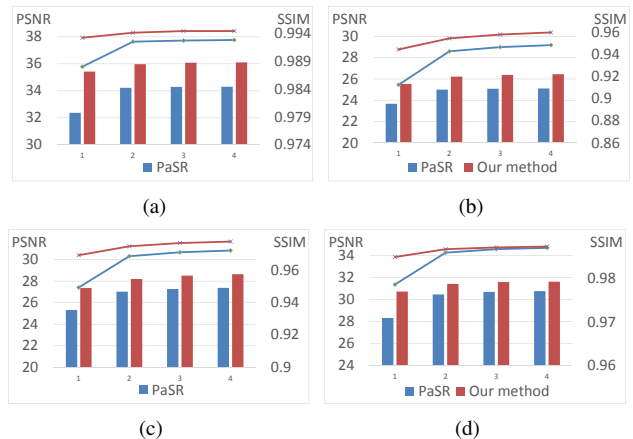


Fig. 11. PSNR and SSIM of PaSR and our method after different numbers of iterations: (a) maria; (b) still life; (c) couple; (d) tarot

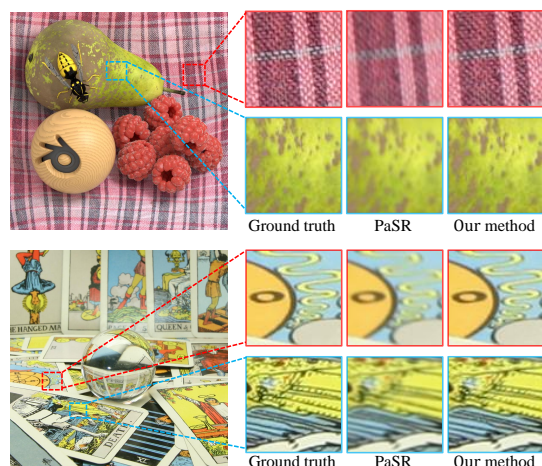


Fig. 12. Comparison of super-resolution results obtained on the data sets still life and tarot.

Different numbers of side-view cameras We evaluate the relationship between the quality of the estimated depth map and the number of side-view cameras using one light field captured by our system. We consider views captured using 1, 2, 4 and 8 side-view cameras, where the selected cameras provide a) only the right side view, b) the left and right side views, c) 4-adjacent views, and d) 8-adjacent views, respectively. As Fig.14 shows, when more side-cameras are used, the quality of the depth map increases; in particular, the depth map generated using 4 or 8 cameras is far superior to that generated using 1 or 2 cameras. In addition, more side-view cameras will provide more angular information, which will provide more help when synthesizing a dense light field.

Dense light field rendering results We synthesize a dense light field in between the captured views using the image from the central high-resolution view and the disparity map obtained via our iPADS method. First, we synthesize a dense light field with 60 new views, from the central view to the right horizontal side view; the EPIs of two different rows are shown in Fig.15. From the clearly structured EPI results, we can conclude that the synthesized light field images are of high quality and have view-consistent disparities. Zoomed-in patches of the yellow Minion's eye from 3 of the 9 views are shown in Fig.16. To illustrate the parallax, mixed images constructed using different color channels from different views are shown and the view variation is obvious. The result of refocusing two of the light fields captured by our system are shown in Fig.18. As shown by the zoomed-in patches, which are focused at three different depths, the depths of field of

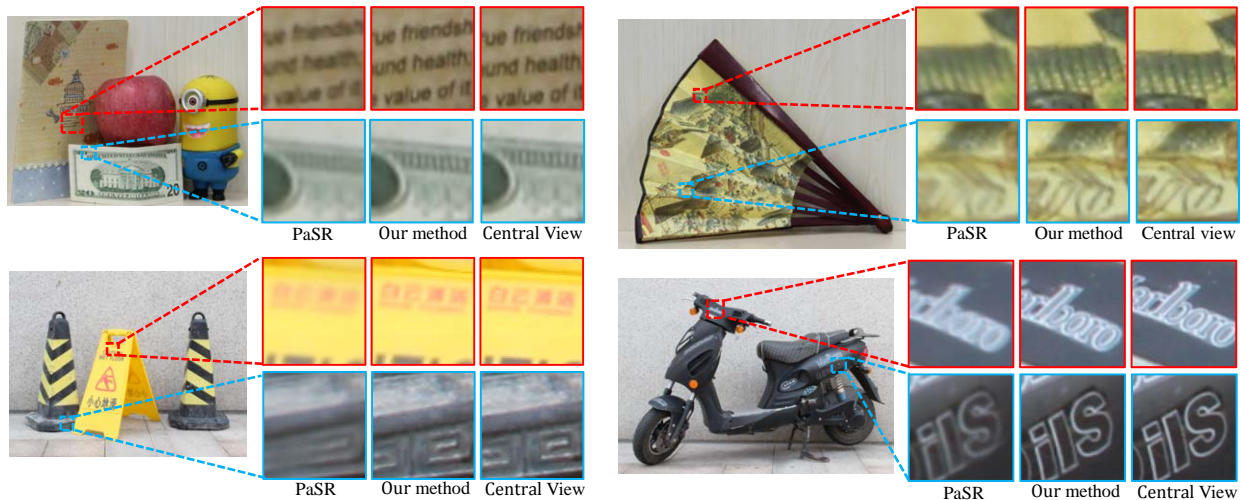


Fig. 13. Comparison of super-resolution results obtained on real data captured by using our light field capture hardware.

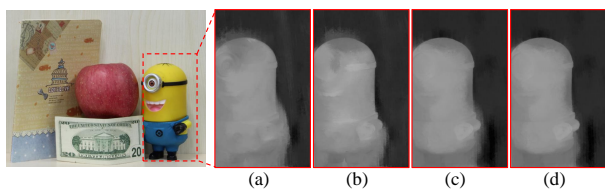


Fig. 14. Depth map estimation results obtained using our method with different numbers of side views: (a) right side view only; (b) left and right side views; (c) 4-adjacent views; (d) 8-adjacent views.

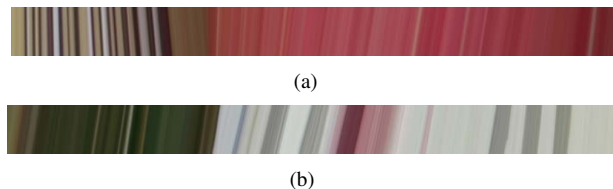


Fig. 15. Two EPIs of a rendered dense light field.

the synthesized images become shallower.

We also compare our dense light field rendering results with those of the traditional DIBR method [39] in the supplemental video. Note that the depth information used in the DIBR method is computed using our proposed pipeline. It can be seen that our proposed iPADS algorithm achieves smooth and coherent angular rendering quality, whereas the DIBR video suffers from jittering. This advantage comes from the phase-based rendering strategy that is naturally integrated into our proposed iPADS framework. More results for dense light field synthesis and its applications are provided in the supplemental materials.

Defocus blur in DSLR images Our algorithm still succeeds when the depth of field of the DSLR camera is quite different than that of the side-view cameras. Fig.17 shows an example in which defocus blur is present in the DSLR image. Although the texture of the image is not clear, the disparity refinement after DAPS yields plausible results, and the blur is coherently propagated to other views in the light field.

Computational efficiency In each iteration, our proposed algorithm performs patch-based super-resolution (step 1), multi-view depth estimation (step 2), depth-aided phase-based synthesis (step 3), optical-flow-based warping (step 4), and a patch database update (step 5). The run times for each step in each iteration are as follows: step 1, 3min, step 2, 5min, step 3, 15min, step 4, 1min, step 5, 12min. The algorithm was implemented in C++

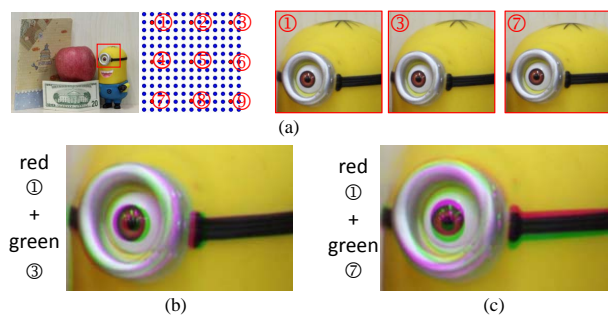


Fig. 16. (a) Three selected views from the synthesizing dense light field. The red circles mark the the selected view positions on the whole light field; (b) mixed image with red and blue channels from view 1, and green channel from view 3; (c) mixed image with red and blue channels from view 1, and green channel from view 7.

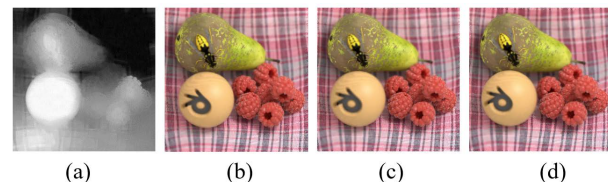


Fig. 17. Results for a case in which defocus blur is present in the DSLR image: (a) DAPS-refined disparity map; (b) and (d) final super-resolution results for the top left and bottom right side views, respectively; (c) captured center-view image. Please zoom in to see the blur.

without optimization on an Intel i7 fourth-generation processor with 32GB of RAM. The algorithm is very suitable for GPU parallel computing (especially in steps 1, 3 and 5), which will significantly speed up the algorithm.

8 CONCLUSION AND DISCUSSION

In this paper, we presented prototype of a portable lens attachment that can take advantage of the popularity of DSLR cameras to capture and reconstruct light fields with high spatial and angular resolution. The data captured by this prototype configuration consist of a high-resolution center-view image and 8 low-resolution side-view images. Based on the captured data, we developed a method called iterative Patch- And Depth-based Synthesis (iPADS) for achieving $8\times$ super-resolution of these low-quality side views and the synthesis of arbitrary in-between views.

Our proposed method has several limitations. First, its performance will be compromised in the case of non-Lambertian



Fig. 18. Refocusing of light fields captured using our system.

scenes or serious occlusions or defocus blur. On the hardware side, our current setup still requires camera calibration. In commercial manufacturing, a mechanically designed locking module can be added for mounting the attachment on the lens, which will enable precise positioning and allow the calibration step to be omitted. Note that the intrinsic parameters of the DSLR camera can be read from the image header, and the intrinsic and extrinsic parameters of side-view cameras on the attachment can be fixed and pre-calibrated. Another limitation is related to the DAPS step. As shown in Fig. 5, some errors are reproduced by the texture here, which are predominantly introduced during the phase-based synthesize (DAPS) step because of its ability to capture detailed disparity information at the cost of over-sensitivity to texture.

REFERENCES

- [1] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. H. Gross, "Scene reconstruction from high spatio-angular resolution light fields," *ACM Trans. Graph.*, vol. 32, no. 4, p. 73, 2013.
- [2] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, "Compressive light field photography using overcomplete dictionaries and optimized projections," *ACM Trans. Graph.*, vol. 32, no. 4, Jul. 2013.
- [3] A. Isaksen, L. McMillan, and S. J. Gortler, "Dynamically reparameterized light fields," in *SIGGRAPH '00*, 2000.
- [4] G. Wetzstein, D. Lanman, M. Hirsch, W. Heidrich, and R. Raskar, "Compressive light field displays," *CG&A*, vol. 32, no. 5, pp. 6–11, 2012.
- [5] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu, "Saliency detection on light field," in *CVPR*, 2014, pp. 2806–2813.
- [6] C.-K. Liang, T.-H. Lin, B.-Y. Wong, C. Liu, and H. H. Chen, "Programmable aperture photography: Multiplexed light field acquisition," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 55:1–55:10, Aug. 2008.
- [7] A. Manakov, J. F. Restrepo, O. Klehm, R. Hegedüs, E. Eisemann, H.-P. Seidel, and I. Ihrke, "A reconfigurable camera add-on for high dynamic range, multi-spectral, polarization, and light-field imaging," *ACM Trans. Graph.*, vol. 32, no. 4, Jul. 2013.
- [8] V. Boominathan, K. Mitra, and A. Veeraraghavan, "Improving resolution and depth-of-field of light field cameras using a hybrid imaging system," in *ICCP*, May 2014.
- [9] M. Levoy and P. Hanrahan, "Light field rendering," in *SIGGRAPH '96*. ACM, 1996, pp. 31–42.
- [10] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *SIGGRAPH '96*. ACM, 1996.
- [11] S. Wanner and B. Goldluecke, "Globally consistent depth labeling of 4D lightfields," in *CVPR*, 2012.
- [12] C. Chen, H. Lin, Z. Yu, S. B. Kang, and J. Yu, "Light field stereo matching using bilateral statistics of surface cameras," in *CVPR*, 2014, pp. 1518–1525.
- [13] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light field microscopy," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 924–934, 2006.
- [14] R. Prevedel, Y.-G. Yoon, M. Hoffmann, N. Pak, G. Wetzstein, S. Kato, T. Schrödel, R. Raskar, M. Zimmer, E. S. Boyden *et al.*, "Simultaneous whole-animal 3d imaging of neuronal activity using light-field microscopy," *Nature methods*, 2014.
- [15] M. Levoy, B. Chen, V. Vaish, M. Horowitz, I. McDowall, and M. Bolas, "Synthetic aperture confocal imaging," in *ACM Transactions on Graphics (TOG)*, vol. 23, no. 3, 2004, pp. 825–834.
- [16] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 765–776, Jul. 2005.
- [17] M. Ziegler, A. Engelhardt, S. Müller, J. Keinert, F. Zilly, S. Foessel, and K. Schmid, "Multi-camera system for depth based visual effects and compositing," in *Proceedings of the 12th European Conference on Visual Media Production*. ACM, 2015, p. 3.
- [18] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *CSTR*, vol. 2, 2005.
- [19] T. Georgiev, "New results on the plenoptic 2.0 camera," in *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on*, Nov 2009.
- [20] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Trans. Graph.*, vol. 26, no. 3, Jul. 2007.
- [21] S. Tambe, A. Veeraraghavan, and A. Agrawal, "Towards motion aware light field video for dynamic scenes," in *ICCV*, 2013, pp. 1009–1016.
- [22] K. Venkataraman, D. Lelescu, J. Duparré, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar, "Picam: An ultra-thin high performance monolithic camera array," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6, p. 166, 2013.
- [23] T.-C. Wang, M. Srikanth, and R. Ramamoorthi, "Depth from semi-calibrated stereo and defocus," in *CVPR*, 2016.
- [24] Raytrix, "3d light field camera technology, <http://www.raytrix.de/>."
- [25] T. Bishop, S. Zanetti, and P. Favaro, "Light field superresolution," in *ICCP*, April 2009, pp. 1–9.
- [26] S. Pujades, F. Devernay, and B. Goldluecke, "Bayesian view synthesis and image-based rendering principles," in *CVPR*, 2014, pp. 3906–3913.
- [27] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *PAMI*, vol. 36, no. 3, pp. 606–619, Mar. 2014.
- [28] P. E. Debevec, C. J. Taylor, and J. Malik, "Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach," in *SIGGRAPH '96*.
- [29] C. Buehler, M. Bosse, L. McMillan, S. Gortler, and M. Cohen, "Unstructured lumigraph rendering," in *SIGGRAPH '01*, pp. 425–432.
- [30] P. Didyk, P. Sitthi-Amorn, W. Freeman, F. Durand, and W. Matusik, "Joint view expansion and filtering for automultiscopic 3d displays," *ACM Trans. Graph.*, vol. 32, no. 6, Nov. 2013.
- [31] 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 (TOG)*, vol. 34, no. 1, p. 12, 2014.
- [32] Z. Zhang, Y. Liu, and Q. Dai, "Light field from micro-baseline image pair," in *CVPR*, 2015, pp. 1–8.
- [33] "Camera calibration toolbox," http://www.vision.caltech.edu/bouguetj/calib_doc/.
- [34] "Sba," <http://users.ics.forth.gr/lourakis/sba/>.
- [35] H. Chang, D. yan Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *CVPR*, 2004.
- [36] H. Shum and S. B. Kang, "Review of image-based rendering techniques," in *VCIP*, 2000, pp. 2–13.
- [37] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," in *CVPR*, 2011, pp. 3017–3024.
- [38] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *ICIP*, vol. 3, 1995, pp. 3444–3444.
- [39] C. Fehn, "Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv," in *Electronic Imaging 2004*, 2004, pp. 93–104.
- [40] C. Liu, "Beyond pixels: exploring new representations and applications for motion analysis," Ph.D. dissertation, 2009.
- [41] Stanford, "The (new) stanford light field archive, <http://lightfield.stanford.edu/>," 2008.
- [42] S. Wanner, S. Meister, and B. Goldluecke, "Datasets and benchmarks for densely sampled 4d light fields," 2013.
- [43] 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*, vol. 13, no. 4, pp. 600–612, 2004.