Stereoscopic Image Generation from Light Field with Disparity Scaling and Super-Resolution

Tao Yan Jianbo Jiao Wenxi Liu Rynson W.H. Lau

***Abstract*—In this paper, we propose a novel method to generate stereoscopic images from light-field images with the intended depth range and simultaneously perform image super-resolution. Subject to the small baseline of neighboring subaperture views and low spatial resolution of light-field images captured using compact commercial light-field cameras, the disparity range of any two subaperture views is usually very small. We propose a method to control the disparity range of the target stereoscopic images with linear or nonlinear disparity scaling and properly resolve the disocclusion problem with the aid of a smooth energy term previously used for texture synthesis. The left and right views of the target stereoscopic image are simultaneously generated by a unified optimization framework, which preserves content coherence between the left and right views by a coherence energy term. The disparity range of the target stereoscopic image can be larger than that of the input light field image. This benefits many light field image-based applications, e.g., displaying light field images on various stereo display devices and generating stereoscopic panoramic images from a light field image montage. An extensive experimental evaluation demonstrates the effectiveness of our method.**

***Index Terms*—Light field image processing, stereoscopic image synthesis, disparity scaling, super-resolution.**

I. Introduction

THE light field technology has gained attention from both academia and industry in recent years. The standard light field camera model [1] relies on the basic principle of placing a micro lens array (MLA) in front of an image acquisition device. A light field camera can capture rich 3D information from a scene by acquiring both spatial and angular light rays to generate a light field image. A light field image can be decoded into a regular array of subaperture views (i.e., multi-perspective imaging of the same scene) [1] [2]. Light field images can be used to facilitate depth map computing, 3D reconstruction, image refocusing, view interpolation, and perspective modification [3].

Recently, 3D stereoscopic images and videos have been widely used on 3D displays in virtual/augmented reality and robotics. A stereoscopic image consists of a pair of images captured from two different viewpoints of the same scene with a specific camera baseline. However, most stereoscopic images suffer from accommodation-convergence conflict and visual-uncomfortable disparity ranges [4] [5] [6]. Such problems can be eliminated by adopting disparity scaling [4] or perspective modification [7]. A typical solution is to change the iter-axial distance of the stereo cameras and/or the convergence point of the optical axis. However, as this solution modifies the depth perception globally, it provides limited control over the local disparity distribution. When this method is used to compress the disparity range of a stereoscopic image, it often results in over-flattening, with most detail changes lost.

Most previous methods related to stereoscopic 3D content creation and manipulation either employ depth-image-based rendering to achieve the intended binocular parallax and image inpainting to fill in the disoccluded monocular regions or adopt image mesh warping to non-uniformly scale the left and right views of the stereoscopic image. The first category of methods usually take a single view, together with an accurate disparity map [8] [9] [10] [11] or manually drawn disparity map [12] [13], as input. The second category of methods always take a stereoscopic image pair as input [4] [14] [7]. However, it is non-trivial to extend these techniques to utilize more than two views to improve the results in the case whereby more views are available in applications such as disparity/depth manipulation based on image warping [4] [14] [7]. On the other hand, the multi-view property of light field images makes it possible to exploit substantially more information to generate high-quality stereoscopic images. To address the aforementioned problems, we propose a novel method for stereoscopic image generation and disparity manipulation from arbitrary viewpoints from light field images.

Compact commercial light field cameras, e.g., those by Lytro [15] and RayTrix [16], usually use a single 2D sensor to multiplex spatial and angular information. Such a setup typically suffers from two problems. First, the setup has either low spatial resolution or low angular resolution. Second, the neighboring subaperture views of a light field image usually have a very small baseline. In recent years, some super-resolution methods have been proposed to address the first problem [17] [18] [19] [20]. For the second problem, the disparity range of neighboring subaperture views is very small (typically in the range of ± 1 pixel [21]) due to the small baseline between subaperture views. Hence, if we generate stereoscopic images directly from light field images, the stereoscopic images tend to have a very small disparity range. To enhance the stereoscopic 3D perception of the target stereoscopic images generated from light field images, their disparity range needs to be carefully scaled during the stereoscopic image generation process. Simultaneously, the resolution of

Tao Yan is with the Jiangsu Key Laboratory of Media Design and Software Technology, Jiangnan University, China. E-mail: yantao.ustc@gmail.com

Jianbo Jiao is with the Department of Engineering Science, University of Oxford, United Kingdom. E-mail: jiaojianbo.i@gmail.com

Wenxi Liu is with the College of Mathematics and Computer Science, Fuzhou University, China. E-mail: wenxi.liu@hotmail.com

Rynson W.H. Lau is with the Department of Computer Science, City University of Hong Kong, Hong Kong. E-mail: rynson.lau@cityu.edu.hk

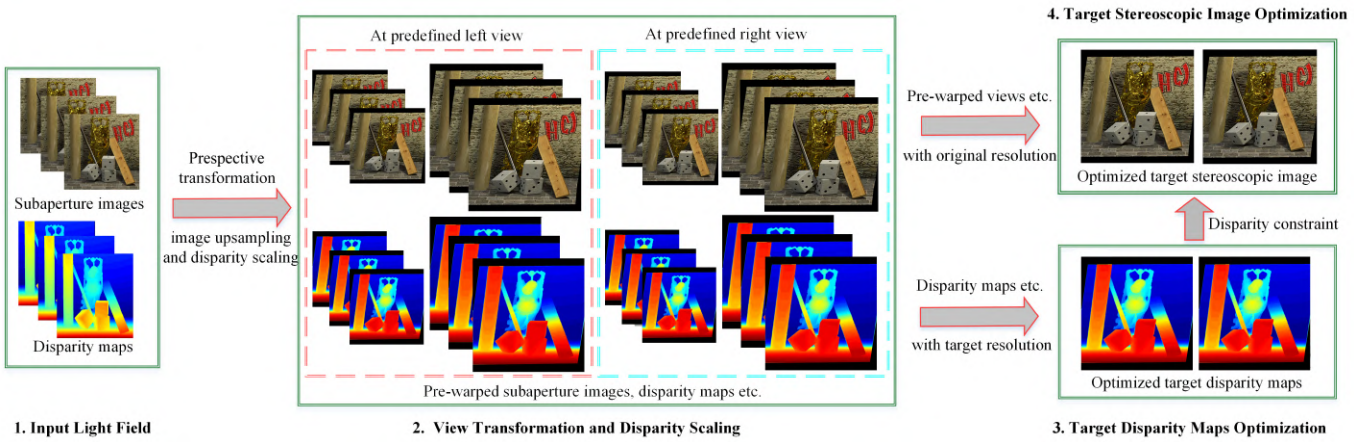


Fig. 1: Overview of our method – stereoscopic image generation from a light field. The input of our method is all the subaperture views of a light field image and its corresponding disparity maps. Our method consists of three main steps. First, our method performs the depth image based perspective projection to warp all subaperture views to target stereo viewpoints with the original and target resolution, respectively. Second, our method optimizes the desired disparity map for the target stereoscopic image pair. Finally, both the perspective-transformed subaperture views and the optimized disparity maps are utilized to generate the target stereoscopic image pair.

the target stereoscopic images also needs to be increased to achieve a higher quality viewing experience.

In this paper, we synthesize stereoscopic images from low-resolution subaperture views of light field images with a specially designed multi-label optimization framework (Fig. 1). After a global optimization, the color of each pixel in the target stereoscopic image pair is determined by the pixel selected from a local search region of the pre-warped subaperture views. Our proposed method is able to synthesize stereoscopic images with the intended disparity constraints and in high resolution, without quality degradation.

The main contributions of this paper include:

1) We propose a novel method to generate stereoscopic images with the intended disparity constraints from light field images. The disparity range of the target stereoscopic images can be larger than that of the light field images.
2) We propose an advisable smooth energy cost term to properly handle the disocclusion problem accompanying disparity scaling and view modification, as well as a coherence energy term to preserve the content coherence between the target left and right views.
3) Our proposed method synthesizes high-resolution output stereoscopic images without introducing noticeable distortion or blurring artifacts.

The remainder of this paper is organized as follows. Sec. II summarizes relevant existing work, followed by some relevant background information on stereoscopic and light field images in Sec. III. Sec. IV presents the details of our method for generating stereoscopic images with the intended disparity range and high resolution. Sec. V presents experimental results and evaluation. Finally, in Sec. VI we conclude our work and discuss possible future work.

II. Related Work

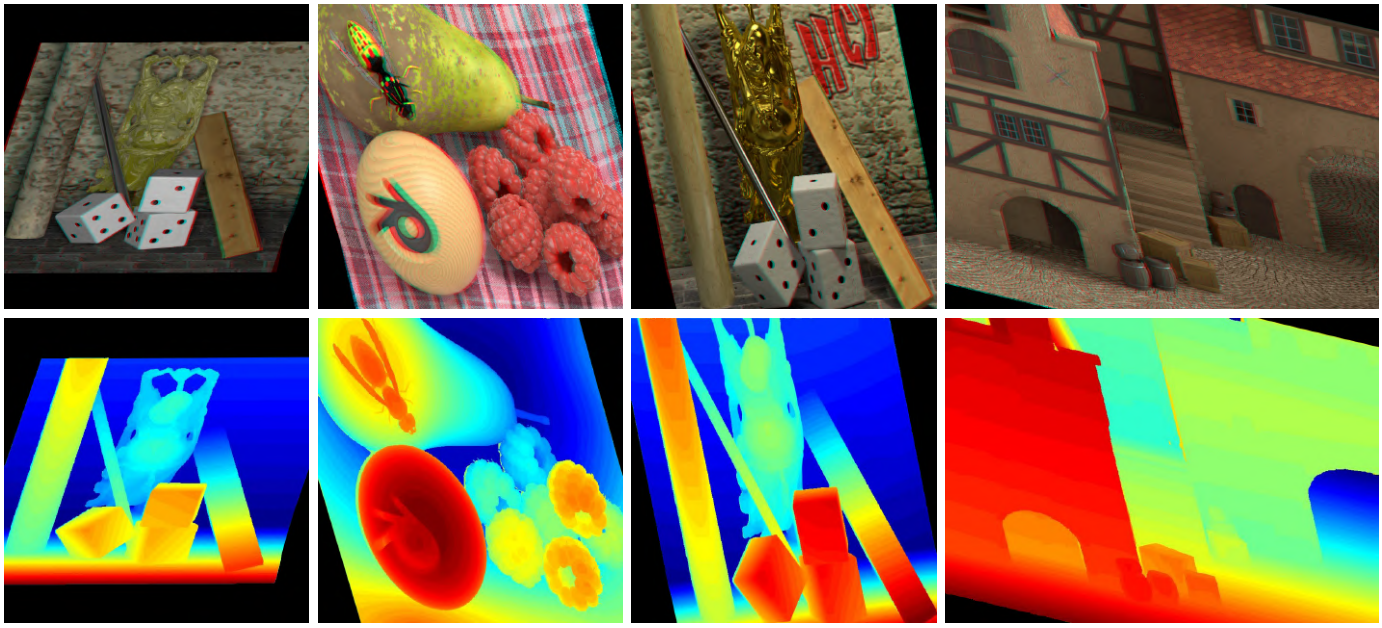
Many studies on light field images have been conducted. In this section, we mainly review recent works related to our work, including novel view synthesis, stereoscopic image generation from light field images, and stereoscopic disparity/depth scaling and perspective manipulation.

A. Novel View Synthesis and Spatial Super-resolution from Light Field Images

Existing methods on view inter-/extrapolation of light field images can be classified into two main categories: total variation based methods [22] [23] and deep convolution neural network (DCNN) based methods [17] [24] [25] [26].

Wanner et al. [22] proposed a framework for light field analysis based on the total variation. This method robustly computes disparity maps and synthesizes novel super-resolution views from light field images. The disparity map computation contains three steps: local depth labeling in the Epipolar-Plane Image (EPI) space utilizing a structure tensor, consistent EPI depth labeling, and depth integration from the horizontal and vertical EPI slices. The authors proposed a variational model to synthesize novel super-resolved views, which works at the sub-pixel level and is efficiently accelerated by GPU card. This method is a state-of-the-art method for estimating disparity maps and synthesizing novel super-resolution views from light field images. However, this method does not perform very well for view extrapolation, especially given the disocclusion problem, due to the insufficient constraints on the total variation. The performance of this method is also readily degraded by inaccurate disparity maps.

Recent representative works based on deep learning [17] [24] can synthesize novel views with super-resolution both in the spatial and angular domains. The method in [17] adopts a data-driven learning approach to up-sample the spatial resolution as well as the angular resolution of a light field



(a) buddha (b) stillLife (c) buddha2 (d) medieval

Fig. 2: Stereoscopic image generation from arbitrary viewpoints. The first row shows stereoscopic images with a depth image based perspective projection effect, produced by our method from light field images. The second row shows the corresponding disparity maps for the left views of the target stereo image pairs. From left to right: (a) shows the stereoscopic image generated from stereo viewpoints $(4, 0)$ and $(4, 8)$ and with the rotation parameters ($\alpha = 0.002\pi$, $\beta = 0$, $\gamma = 0.002\pi$). (b) shows the stereoscopic image generated from stereo viewpoints $(4, 0)$ and $(4, 8)$ and with the rotation parameters ($\alpha = -0.002\pi$, $\beta = 0.003\pi$, $\gamma = -0.005\pi$). (c) shows the stereoscopic image generated from stereo viewpoints $(4, -2)$ and $(4, 6)$, and with the rotation parameters ($\alpha = 0.002\pi$, $\beta = -0.001\pi$, $\gamma = 0.002\pi$). (d) shows the stereoscopic image generated from stereo viewpoints $(4, -3)$ and $(4, 11)$ and with the rotation parameters ($\alpha = -0.002\pi$, $\beta = 0.003\pi$, $\gamma = -0.005\pi$). Intermediate results, such as warped subaperture views with black holes, can be found in the supplemental.

image. The method first increases the spatial resolution of each subaperture view and enhances the details of the image content using a spatial super-resolution network with four layers. The method then generates novel views between subaperture views using an angular super-resolution network with four layers. The two networks are trained independently, and then fine-tuned via end-to-end training. The method in [24] is an extension of [17]. That method improves both the efficiency of the training process and the quality of the angular super-resolution results by using weight sharing in an angular super-resolution network. However, these two methods only use two adjacent subaperture views for the intermediate view synthesis, therein underutilizing all the information provided by the light field image. In addition, those methods could only achieve a fixed super-resolution rate of $\times 2$.

Flynn et al. [25] designed a DCNN-based method consisting of a selection tower and a color tower to generate novel views from multiview stereo images. The selection tower predicts the approximate depth of each pixel in the output image and determines which source image pixels could be used as candidate pixels to generate the output pixels. The color tower utilizes all relevant source image pixels to produce a color for that output pixel. This deep architecture is trained on a large number of posed image sets, such as Google's Street View image database and the KITTI dataset. With this method, pixels from the neighboring views of a scene are presented to the network, which then directly produces the pixels for the novel view. This method can be easily modified to synthesize novel views from light field images.

Kalantari et al. [26] proposed a representative DCNN-based method to synthesize novel views from light field images. This method has two main parts; each is a simple DCNN model with four layers. First, this method synthesizes the target disparity map for a specified novel view with one DCNN model. Second, the method takes the pre-warped subaperture views and the target disparity map from the first step as input to synthesize the target view with another DCNN model. However, this type of DCNN-based models for view inter-/extrapolation cannot address the global/local disparity scaling problem for enhancing the stereoscopic 3D perception of the target stereoscopic image, because such a problem is ill-proposed and there is no ground truth available for training.

Wu et al. [27] proposed a DCNN-based method for super-resolution in the angular space of light field images. That method reconstructs a high angular light field on EPI. To avoid ghosting effects caused by the information asymmetry, the low-frequency spatial information of the EPI is extracted via EPI blur and then used to recover the angular detail. Later, the non-blind deblur operation is used to restore the spatial detail suppressed by the EPI blur. However, this method cannot be utilized for inter-/extrapolation from arbitrary viewpoints and

super resolution.

Vagharshakyan et al. [28] proposed a method for reconstructing a high-angular-resolution light field image from a small number of rectified multiview images taken with a wide baseline. This method adopts a sparse representation of the underlying EPIs in the shearlet domain and employs an iterative regularized reconstruction. Similarly, this method cannot be utilized for inter-/extrapolation from arbitrary viewpoints and spatial super-resolution.

Rossi et al. [18] proposed a novel light field super-resolution method based on graph-cut optimization. It adopts a multi-frame-like super-resolution approach, where the complementary information in different subaperture views is used to increase the spatial resolution of the light field image. The method utilizes a graph regularizer to enforce the light field structure based on non-local self-similarity, avoiding the challenging disparity estimation. The optimization is based on the graph cut method [29] [30]. The method is only tested on small-sized light field images, which are down-sampled to half the spatial resolution of the subaperture views. The method also does not consider the disparity scaling function.

Two methods, [19] and [20], utilize a hybrid stereo imaging system consisting of a light field camera and a traditional digital single-lens reflex (DSLR) camera for light field image super-resolution. Method [19] utilizes a patch-based algorithm to super-resolve the low spatial resolution of the subaperture views by utilizing the high-resolution patches captured using a high-resolution SLR camera. Method [20] warps a high-resolution image captured by a DSLR camera onto each bicubically upsampled light field subaperture view with the optical flow estimated between the high-resolution 2D image and subaperture views of the light field images. The warped high-resolution and upsampled subaperture views are fused to generate the high-resolution subaperture views by a wavelet-based approach and alpha blending.

B. Stereoscopic Images Generating from Light Field Images

Recently, based on graph cut and convex variational optimization, several works [31] [32] [23] have been proposed for stereoscopic image generation from light field images. Kim et al. [31] proposed a method for producing stereoscopic images with the expected disparity range distribution by utilizing a 4D graph cut method on the Epipolar-Plane Image(EPI) space. Although high-quality stereoscopic images can be generated, the disparity range is strictly limited by the disparity range of the input light field images. In addition, it only supports depth adjustment for selected objects instead of global adjustment.

Kim et al. [32] proposed another method that adopts multi-perspective imaging for stereoscopic image synthesis from light field images. Taking a light field image and a manually drawn disparity map as input, the method synthesizes the target view by selecting light rays that satisfy the given disparity constraints. More concretely, the method chooses a subaperture view as one view of the target stereoscopic image, and the unknown second view is generated by solving a multi-label optimization problem based on convex variation to determine which subaperture view each pixel should be taken from. Essentially, this method performs view interpolation, and the disparity range of the target stereoscopic image is strictly constrained by the input light field image. Our method can perform more flexible linear/nonlinear disparity scaling and view inter-/extrapolation.

Zhang et al. [23] proposed a method to synthesize stereoscopic images with the intended disparity constraints from a light field image. By specifying one subaperture view as one view of the target stereoscopic image, that method adopts variational optimization to synthesize another view of the target stereoscopic image by using weighted view inter-/extrapolation. The authors adopted linear, non-linear, and artistic disparity scaling on the original disparity range. A fast optimization method based on convex total variational is proposed to speed up the computation for target view. The method is an extension of [22], and it inherits the limitation of [22] . We observe from the experimental results that this method always assigns larger disoccluded regions with the same color (see Fig. 5 in [23]). This means that the method cannot synthesize meaningful textures for the disoccluded regions. In addition, noise, i.e., incorrectly optimized color values, can always be found in new synthesized images, which means stereoscopic images produced by this method can easily be disturbed by incorrect disparity maps.

The most closely related works to ours are [32] and [23]. Method [32] supports a limited range of disparity scaling restricted by the original disparity range of the input light field image. Method [23] cannot solve the disocclusion problem well and may leave black holes in the synthesized images.

C. Stereoscopic Image Depth Manipulation and Editing

Lang et al. [4] proposed a nonlinear disparity scaling method for stereoscopic images based on image mesh warping, which treats the image domain as a continuous 2D grid. Thus, it is unable to address occlusion and disocclusion problems. In addition, image warping can easily introduce content distortion on image features such as edges and lines. For stereoscopic images, mesh warping methods can easily distort 3D features and scene structures. Yan et al. [14] proposed a method based on image warping to scale the disparity/depth range of stereoscopic images, while simultaneously preserving important 3D scene features/structures. Du et al. [7] proposed the first method for the perspective view manipulation of stereoscopic images. This method also adopts image warping based on quad face mesh warping. In contrast to view inter-/extrapolation, perspective view manipulation is another type of stereoscopic image post-processing and editing method. Yan et al. [33] proposed a consistent stereo image editing framework that extended the shift map strategy for regular 2D image editing [34] to stereoscopic image editing. This method supports disparity/depth scaling, specified object depth adjustment and non-homogeneous stereoscopic image resizing. However, it cannot deal with light field image editing tasks that need to handle multi-perspective views.

We propose a novel method to generate stereoscopic images with a large variation in disparity range scaling. Our method takes a light field image and the corresponding disparity maps

as input. It simultaneously generates the left and right views of the target stereoscopic image to satisfy the intended disparity scaling constraints. It can also properly handle the occlusion and disocclusion problems, and the left view and right view coherence problem. Further, our method can generate super-resolved stereoscopic images.

III. Background

A. Stereoscopic Images

A stereoscopic image pair consists of two views: left view and right view. Suppose that the stereo camera follows a convergence capturing model. Then, disparity value of two corresponding pixels projected by a 3D point onto a stereoscopic image pair is mainly determined by the focal length, the baseline of the stereo cameras and the focal plane distance [1], which are typical defined as follows:

$$d = \frac{Bf}{z_d} - \frac{Bf}{z_o}, \tag{1}$$

where z_o is the distance from the focusing plane to the stereo camera. z_d is the depth of the 3D point. f is the camera focal length. B is the baseline, i.e., the distance between the optical centers of the left and right cameras; $\frac{Bf}{z_o}$ is the shift (i.e., offset) between the left and right views.

The 3D stereoscopic effectiveness of the stereoscopic image as perceived by the viewers is related to the disparity range and the camera baseline [14], defined as follows:

$$z_p = \frac{et}{e - s}, \tag{2}$$

where z_p is the viewer's perceived depth value. e is the distance between the viewer's two pupils. t is the distance between the viewer and the target display screen. s is the parallax value of the corresponding pixels shown on the target screen and is defined as $s = \beta \times d$. β is a constant parameter related to the physical pixel size of the target display.

B. Light Field Images

A light field image consists of multiple subaperture views. The light field image not only records the integrated color intensity of each pixel in the image sensor but also the direction of each ray that passes through the main lens and microlens array. The number of microlenses in the microlens array determines the spatial resolution of the light field images [1] [2]. The resolution of each microlens image determines the number of subaperture views, i.e., the angular resolution. For standard plenoptic cameras, because the spatial and angular resolutions share the same image sensor with a limited resolution, both the spatial and angular resolutions of the light field image cannot be very large simultaneously. For a light field image, the stable and high-quality disparity/depth maps can be restored by analyzing the EPI [35] [36] [37] [22], utilizing the phase shift theorem to construct the relationship between subaperture views [21], or exploiting the features of optical stacks [38] [39]. Most existing methods only output a single disparity map for the central subaperture view that can

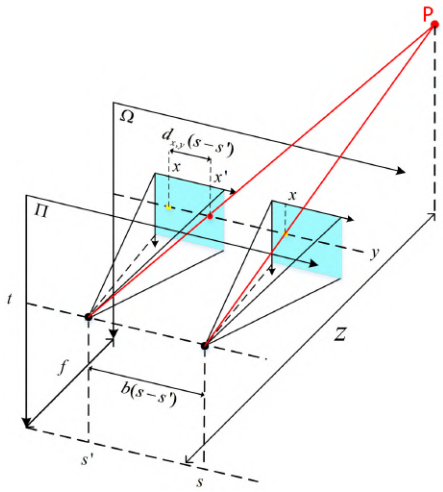


Fig. 3: Illustration of light field imaging. Two light rays (red lines) are from the same 3D point P through the image plane Ω and view plane Π.

then be used to infer the disparity maps for other subaperture views through the following relationship [22] [23] [21] [18]:

$$x_j = x_i - d_i(x_i)(c_i - c_j), \tag{3}$$

where c_i is a subaperture view of the light field image. c_j is one of the other views of the light field image. d_i is the disparity map for the subaperture view i. x_i is the pixel coordinate in the subaperture view i. x_j is a pixel in subaperture view c_j. If x_j is not occluded, x_i and x_j should be the projections from the same 3D point. Our method can utilize this relationship to project a subset of subaperture views of the input light field image to the target view for view inter-/extrapolation.

Fig. 3 shows the above relationship between two subaperture views, and the corresponding image and viewpoint planes. (s, t) and (s', t) represent two viewpoints in Π, corresponding to two subapertures views of the light field. The two pyramids represent the projection of two pinhole cameras. Each pixel in the projected image of a pinhole camera can be represented by a 4D coordinate that denotes the relative coordinate between the image pixel and the optical center of the subaperture view, e.g., (s, t, x, y) and (s', t, x', y). (x, y) and (x', y) denote the projected points of P in subaperture views (s, t) and (s', t), respectively. The baseline between two neighboring subaperture views with the same vertical coordinate in Π is b. The baseline between two subaperture views (s, t) and (s', t), with the same vertical coordinate, is $b(s-s')$. If the disparity of two projected points of P in two neighboring subaperture views is d_{xy}, the disparity of two projected points of P in subaperture views (s, t) and (s', t) will be $d_{xy}(s - s')$. This relationship is similar for two subaperture views with the same horizontal coordinate in the vertical direction on the view plane Π. The distance between two planes Π and Ω is f. The distance from P to plane Π is Z.

The non-parallel camera model discussed in Sec. III-A is suitable for the subaperture view triangulation model [1] of a standard light field camera. With this standard camera,

the sensor planes of the virtual cameras are coplanar, while their optical axes intersect at a point on the focal plane. The disparity relationship between two subaperture views of a light field image satisfies the relationship defined by Eq. 1. If we want to change the resolution of the target stereoscopic image pair captured at two different viewpoints without changing their baseline, we only need to modify the focal length of the stereo cameras in the capturing stage. The disparity of the light field image will then be linearly scaled with the spatial resolution/focal length of the subaperture views.

IV. Our Method

We propose a method to generate stereoscopic images from light field images with the intended disparity range and high resolution. Our method is based on the Markov Random Field (MRF) framework. We synthesize the target stereoscopic image from the warped subaperture views of the input light field image. For a given pixel in the target stereoscopic image pair, we try to select an optimal pixel from its local search region of the warped subaperture views and assign its color to the given pixel. To label each pixel in the target stereoscopic image pair with corresponding color value, our method addresses a carefully designed discrete multi-labeling optimization problem with the graph-cut based optimization strategy [29] [30].

As illustrated in Fig. 1, our method has three main steps. First, we warp each subaperture view to the specified stereo viewpoint pair for the target stereoscopic image via DIBR (depth image based rendering). At the same time, the disparity range of the target stereoscopic image is tuned by disparity scaling (shifting each pixel in the warped subaperture images). Second, disparity maps for the target stereoscopic image pair are optimized by utilizing the disparity maps corresponding to the warped subaperture views. Third, a global optimization framework is proposed to generate the target stereoscopic image pair based on the optimized disparity maps and warped subaperture views.

A. View Transformation and Disparity Scaling

Given the target stereo viewpoint pair and disparity range, we perform perspective projection to separately transform all the pixels in each subaperture view to the two viewpoints of the target stereoscopic image. Meanwhile, a disparity scaling function is leveraged to shift each pixel again in the warped subaperture views in order to meet the desired disparity scaling requirement. As a result, two warped view stacks, I_L^{lw} and I_R^{lw}, are produced, which consist of the above warped subaperture views for the target stereo viewpoint pair. The warped views in the above stacks are then fused to form the target left and right views.

Because each subaperture view of the input light field image is captured from a slightly different perspective of the scene, transforming multiple subaperture views to a specific viewpoint can leverage more information to generate a high-quality target image, and at the same time resolve occlusion/disocclusion and antialiasing. More importantly, by transforming all subaperture views of the input light field image to a stereo viewpoint pair, the stereoscopic image generation problem can be formulated as a multi-label optimization problem by selecting pixels from a local search region of the warped subaperture view stack to synthesize the target stereoscopic image.

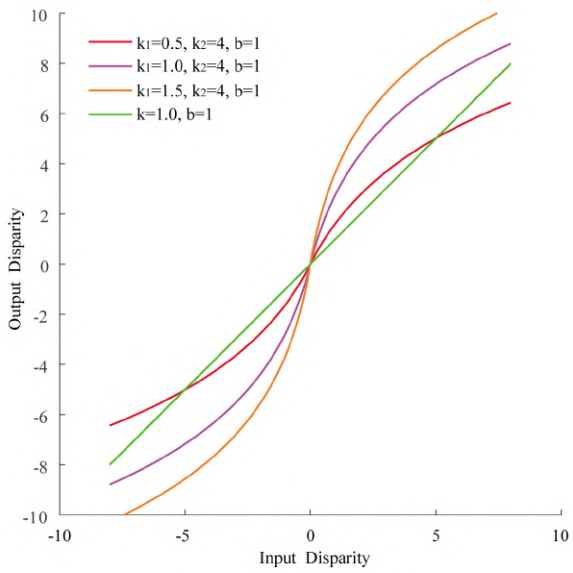


Fig. 4: The disparity scaling function used in our method. Different parameter settings for the nonlinear scaling function (Eq. 5) are shown as red, magenta, and original curves. The green line represents the linear scaling function (Eq. 4) with $k = 1.0$ and $b = 0$.

1) Disparity Scaling: The disparity range of the target stereoscopic image can be modified by a linear f_l or nonlinear f_n scaling function as follows:

$$f_l(d) = k_0 d + b_0, \quad (4)$$

$$f_n(d) = s(d) k_2 \ln(k_1 |d| + b_1), \quad (5)$$

where d is the original disparity value of the target stereoscopic image. k_0, b_0, k_1, k_2 and b_1 are constant parameters for disparity scaling. s is the sign function, and $\ln$ is the natural logarithm.

Since the disparity range of any two adjacent subaperture views is very small and because disparity values can be negative or positive, directly utilizing logarithms as in [4] [40] for the disparity scaling of light field images neither addresses the negative disparity values nor enlarges the disparity range. Thus, we propose a nonlinear disparity scaling function (Eq. 5) to enlarge the disparity range and enhance the local depth contrast of the target stereoscopic image.

In Eq. 5, we set k_2 as the interval between two adjacent subaperture views, i.e., $k_2 = s_R - s_L$, where s_L and s_R are the horizontal coordinates of the target stereo viewpoint pair. Thus, if k_1 is equal to $1/k_2$, the natural logarithm function Eq. 5 directly scales the original disparity range (i.e., the disparity range of the nearest adjacent subaperture views). If k_1 is larger than $1/k_2$, we first enlarge the original disparity range, and then perform nonlinear scaling with the logarithm function to enhance the local depth/disparity contrast for the disparity range of the target stereoscopic image. Finally, the

scaled disparity range is further enlarged with the scaling parameter k_2. Thus, we set $k_2 = s_R - s_L$ and $b_1 = 1$ and allow the user to freely adjust k_1 for a nonlinear disparity scaling function.

In this work, we adopt the nonlinear disparity scaling function to enlarge the disparity range and enhance the local disparity for synthesized stereoscopic images. In addition, we utilize the linear disparity scaling function with $k_0 = 1$ and $b_0 = 0$ to synthesize a novel view for stereoscopic image generation without super-resolution. For the super-resolution of stereoscopic image synthesis, we adopt a linear disparity scaling function with k_0 equaling the upsampling factor and $b_0 = 0$ to scale the original disparity. Fig. 4 illustrates the proposed disparity scaling function.

For each warped subaperture view in the above image stacks, we compute its corresponding mask map (i.e., M_L^{lw} or M_R^{lw}) and disparity map (i.e., D_L^{lw} or D_R^{lw}). A pixel in the warped subaperture view with a valid mask value indicates that its color and disparity are provided by the original subaperture view through our perspective projection and disparity scaling.

If an output high-resolution stereoscopic image is required, two warped subaperture view stacks, I_L^{hw} and I_R^{hw}, are generated at the same resolution as the target stereoscopic image. Note that the resolution and disparity values of the original disparity map should be scaled first before performing view transformation or disparity scaling. (Refer to Sec. III-B for a discussion on disparity linearly scaling with the resolution of the subaperture views). The corresponding mask map (i.e., M_L^{hw} or M_R^{hw}) and disparity map (i.e., D_L^{hw} or D_R^{hw}) are also calculated for each warped subaperture view at high resolution.

2) Perspective Projection: To generate target stereoscopic images, all subaperture views are first projected to the desired stereo viewpoint pair. Because each subaperture view of a light field image possesses the estimated disparity map, we obtain warped subaperture views in the desired viewpoints of the target stereoscopic images based on the DIBR as follows:

$$p' = P(T(p, v), v'), \tag{6}$$

where the function T projects pixel $p = (x, y)$ within a subaperture view defined at the viewpoint $v = (s, t)$ to the 3D coordinate (X, Y, Z) according to

$$X = \frac{Z}{f}(x - x_0 + d(x, y)(s - s_0)), \tag{7}$$

$$Y = \frac{Z}{f}(y - y_0 + d(x, y)(t - t_0)), \tag{8}$$

$$Z = \frac{bf}{d(x, y) + d_s}, \tag{9}$$

where (x_0, y_0) is the spatial coordinate of the central pixel in a subaperture view and (s_0, t_0) is the angular coordinate of the central subaperture view of the input light field image. The 3D coordinate system originates at the viewpoint of the central subaperture view of the input light field image.

The function P projects the 3D coordinate of p, i.e., $[X, Y, Z]^T$, back to the view $v' = (s', t')$ to obtain the pixel $p' = (x', y')$.

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} f & 0 & x_0 & 0 \\ 0 & f & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{10}$$

where $R(\alpha, \beta, \gamma) = R_x(\alpha)R_y(\beta)R_z(\gamma)$ is the rotation matrix for the rotation angles α, β and γ separately rotating about the X, Y and Z axes, respectively. $T = [b(s' - s_0), b(t' - t_0), 0]^T$ indicates the translation vector of the new viewpoint, $v' = (s', t')$, relative to the origin of the 3D coordinate system.

To avoid introducing vertical or other incorrect disparities into the target stereoscopic image, we assume that the angles of rotation about each axis for the two viewpoints of the target stereoscopic image pair take the same value. Some stereoscopic images generated from light field images with perspective projection are shown in Fig. 2.

3) Confidence Map: As there are always noise/errors in the disparity maps for the subaperture views of input light field images, there may be inconsistences between the roughly warped subaperture views. Hence, we propose a metric to measure the confidence of the prior information provided by our perspective projection and disparity scaling steps. Let C_L and C_R denote the confidence maps for the target left and right views. Taking C_L as an example, the definition is:

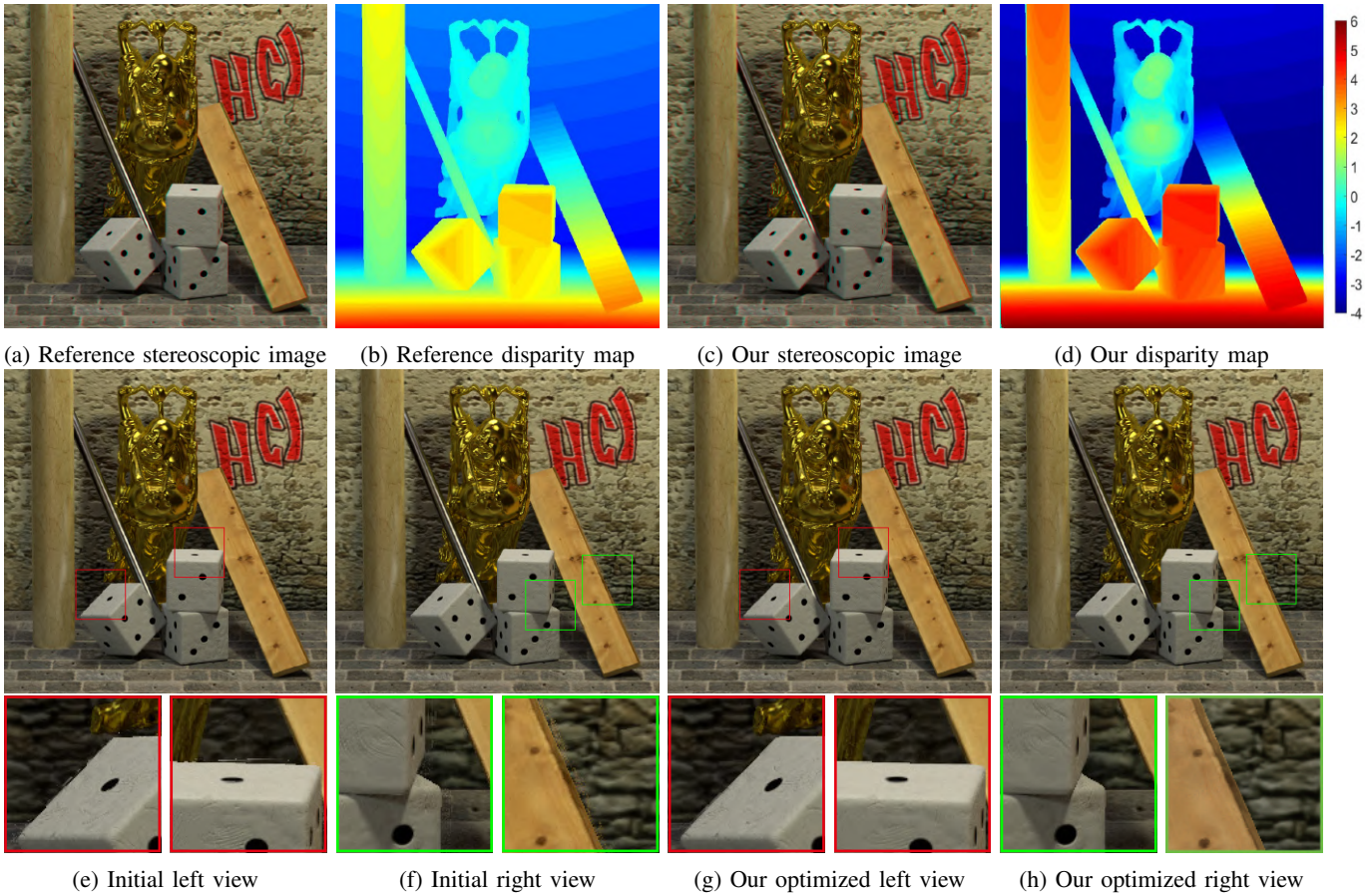
$$C_L = e^{-\omega_d D_e} S, \tag{11}$$

where S is computed as the mean values of the gradient maps of the warped subaperture views (I_L^{lw} or I_R^{lw}). D_e is the standard deviation of the prior color information provided by the warped subaperture views and is calculated from the warped subaperture views (I_L^{hw} and I_R^{hw}). ω_d is a constant weighting parameter and is set to 0.1. The confidence map is normalized to the range of $[0, 1]$.

We observe that the impact of the disparity scaling may vary in different regions of a view. For example, it is not necessary to rigidly shift all the pixels within a homogeneous region, as it seldom introduces noticeable distortion to the corresponding region in the target stereoscopic image pair. In contrast, human eyes are more sensitive to edge pixels with high gradient values, and such pixels should strictly satisfy the disparity scaling constraints. We use these confidence maps to guide our stereoscopic image generation.

B. Disparity Map Optimization for Target Stereoscopic Image

Once the disparity maps for the warped subaperture views are obtained, disparity maps corresponding to the target stereoscopic image pair can be obtained with the following



(a) Reference stereoscopic image (b) Reference disparity map (c) Our stereoscopic image (d) Our disparity map

(e) Initial left view (f) Initial right view (g) Our optimized left view (h) Our optimized right view

Fig. 5: Nonlinear disparity remapping. Top row (left to right): (a) stereoscopic image composed of two subaperture views $(4, 2)$ and $(4, 6)$ of the input light field image, (b) the corresponding disparity map, (c) the stereoscopic image optimized by our method using a nonlinear disparity scaling function (Eq. 5) with parameters set to $k_1 = 0.75$, $k_2 = 4$, $b_1 = 1$, and (d) the corresponding disparity map. Bottom row (from left to right): initial warped (e) left view and (f) right view at the above two fixed viewpoints without using our optimization method, and using our optimization method to generate the (g) left view and (h) right view. This light field image comes from the dataset [41].

optimization:

$$
\begin{aligned}
(D'_L, D'_R) = \arg\min \sum_{y=0}^{M-1} \sum_{x=0}^{N-1} (\\
\omega_{dd} (\sum_{j=0}^{n-1} M^{hw}_{L,j}(x,y) \min(|D'_L(x,y) - D^{hw}_{L,j}(x,y)|^2, \tau_1) \\
+ \sum_{j=0}^{n-1} M^{hw}_{R,j}(x,y) \min(|D'_R(x,y) - D^{hw}_{R,j}(x,y)|^2, \tau_1)) \\
+ \omega_{ds} (\sum_{j=0}^{n-1} \min(|D'_L(x,y) - D'_L(x_b,y_b)|, \tau_2) \\
+ \sum_{j=0}^{n-1} \min(|D'_R(x,y) - D'_R(x_b,y_b)|, \tau_2)) \\
+ \omega_{dc} \min(|D'_L(x,y) - D'_R(x',y)|, \tau_3)),
\end{aligned}
\tag{12}
$$

where $[N, M]$ is the resolution of the target stereoscopic image. n is the number of subaperture views of the input light field image. D'_L and D'_R are the optimized disparity maps for the target stereoscopic image pair. $D^{hw}_{L,j}$ and $D^{hw}_{R,j}$ are the disparity maps for the warped and up-sampled subaperture views $I^{hw}_{L,j}$ and $I^{hw}_{R,j}$, while $M^{hw}_{L,j}$ and $M^{hw}_{R,j}$ are the mask maps for $I^{hw}_{L,j}$ and $I^{hw}_{R,j}$. (x_b, y_b) represents a neighboring pixel of pixel (x, y). Pixels (x, y) and (x', y) represent a corresponding pixel pair in the left and right views, respectively. $\omega_{dd} = gC_{L/R}(x, y)$ represents the weighting parameter for the data cost, which is obtained by multiplying the confidence map by a constant weighting parameter g. ω_{ds} and ω_{dc} are constant weighting parameters for the smooth energy term and coherence energy term. τ_1, τ_2 and τ_3 are constant threshold parameters. The above minimizing function can be solved by a graph-cut-based multi-label optimization method [29].

C. Target Stereoscopic Image Generation

We design an optimization framework to synthesize the target stereoscopic image from a light field image. Each pixel in the left and right views is represented as a node in our constructed graph model. An optimal label is consequently assigned to the pixel. The objective energy function is minimized by a multi-label optimization method [29]. This label of a node

refers to the pixel in the search region whose intensity will be assigned. For instance, a node $p(x,y)$ with label $l(p_i,p_x,p_y)$ in the target left view means that this pixel will be assigned the intensity of pixel $(x/s+p_x, y/s+p_y)$ from the warped subaperture view I^{lw}_{L,p_i}, where s is the upsamping factor.

1) Data Energy Constraint: We define a data energy term to measure the satisfaction of the optimized color and disparity for each pixel with the prior information provided by the warped subaperture views and disparity maps in Sec. IV-A. The final color assigned to a pixel after our target optimization is determined by its optimized label, which refers to the candidate pixel to be assigned in the target search region of the warped view stack. We define our data term as follows:

$$E_{data}(x,y) = \frac{1}{n}\sum_{j=0}^{n-1} M^{lw}_j(x',y')W^T_d(x,y) \min(|V_{p_i}(x',y') - \hat{V}_j(x,y)|^2, T_d), \quad (13)$$

where $(x',y') = (x/s+p_x, y/s+p_y)$. s is the upsamping factor. $V_{p_i}(x',y') = [I^{lw}_{p_i}(x',y'), D^{lw}_{p_i}(x',y'), p_y, p_x]^T$. $\hat{V}_j(x,y) = [I^{hw}_j(x,y), D'(x,y), 0, 0]^T$. $W_d(x,y) = [C(x,y), \omega_d, \omega_l, \omega_l]^T$. M^{lw}_j is the mask map for $I^{lw}_{L,j}$ or $I^{lw}_{R,j}$. $M^{lw}_j(x',y') \in \{0,1\}$ indicates whether pixel (x',y') has valid information provided by the warped subaperture view I^{lw}_j, as discussed in Sec. IV-A. n is the total number of subaperture views. The vector T_d is a constant threshold parameter of our data cost energy term.

The vector (p_i,p_x,p_y) is the label for pixel (x,y) in one view of the left or right view of the output stereoscopic image. This means that pixel (x,y) is assigned the intensity of pixel (p_x,p_y) in the warped subaperture view p_i. $I^{lw}_{p_i}$ is the warped subaperture view I^{lw}_{L,p_i} or I^{lw}_{R,p_i}, and D^{lw} is the corresponding disparity map of $I^{lw}_{p_i}$. I^{hw}_j is the warped and up-sampled subaperture view $I^{hw}_{L,j}$ or $I^{hw}_{R,j}$ in the view stack with super-resolution for the left or right view of the target stereoscopic image. C denotes the confidence map (C_L or C_R) of the target stereoscopic image defined in Eq. 11. D' refers to the above computed disparity map D'_L or D'_R defined in Eq. 12. $D^{lw}_{p_i}$ is the disparity map of the warped subaperture view I^{lw}_{L,p_i} or I^{lw}_{R,p_i}. ω_d and ω_l denote the constant weighting parameters for the disparity and label cost, respectively.

2) Smooth Energy Constraint: We use an undirected edge between two neighboring nodes to represent the penalty of two neighboring pixels taking different labels. Inspired by [42] [34], our smooth cost energy term is defined as follows, in which the gradient is replaced by the disparity:

$$E_{sm}(x,y) = W^T_s(\min(|V_{p_i}(x'_p,y'_p) - V_{q_i}(x'_q-\Delta_x, y'_q-\Delta_y)|, T_{sm}) + \min(|V_{q_i}(x'_q,y'_q) - V_{p_i}(x'_p+\Delta_x, y'_p+\Delta_y), T_s|)), \quad (14)$$

where $(x'_p,y'_p) = (x_p/s+p_x, y_p/s+p_y)$. $(x'_q,y'_q) = (x_q/s+q_x, y_q/s+q_y)$. $V = [I^{lw}, D^{lw}, p_y, p_x]^T$, $W_s = [\omega_{sc}, \omega_{sd}, \omega_{sl}, \omega_{sl}]^T$. I^{lw} and D^{lw} are the warped subaperture view and corresponding disparity map of the input light field image in the original spatial resolution, respectively. p and q are two neighboring pixels. (p_i,p_x,p_y) and (q_i,q_x,q_y) are labels for p and q, respectively. $(\Delta_x,\Delta_y) = (q_x-p_x, q_y-p_y)$. ω_{sc}, ω_{sd} and ω_{sl} are constant weighting parameters. The vector T_s is a threshold parameter of our smooth energy term.

The definition of our smoothness term satisfies the important metric constraint requirement of graph-cut-based multi-label optimization methods [29] [30].

Our smooth energy term can be used to synthesize textures for disoccluded regions, in the same manner that prior image editing works has addressed texture synthesis [43] [34]. In this way, our method can handle the propagation of meaningful textures to disoccluded regions instead of simply filling such regions with a similar color, as in the previous total-variation-based method [23].

3) Coherence Energy Constraint: Assuming that p and q are two different pixels within the left and right views of the target stereoscopic image, if they satisfy the stereo matching relationship, i.e., $p_y = q_y$, $p_x - D_L(p_x,p_y) = q_x$ and $D_L(p_x,p_y) = D_R(q_x,q_y)$, they will form a pixel pair in the target stereoscopic image.

We utilize the following energy term to ensure that the corresponding pixel pair in the synthesized left and right views will have a consistent color and disparity:

$$E_{co}(x,y) = W_c \min(|V_{p_i}(x'_p,y'_p) - V_{q_i}(x'_q,y'_q)|, T_c), \quad (15)$$

where $(x'_p,y'_p) = (x_p/s+p_x, y_p/s+p_y)$. $(x'_q,y'_q) = (x_q/s+q_x, y_q/s+q_y)$. $V = [I^{lw}, D^{lw}]^T$. $W_c = [\omega_{cc}, \omega_{cd}]^T$. p and q are corresponding pixels in the left and right views, respectively. ω_{cc} and ω_{cd} are constant weighting parameters for color and disparity coherence in the target stereoscopic image pair, respectively. The vector T_c is a threshold parameter for our coherence energy term.

Total Energy Function: Finally, we minimize the following energy function to obtain the target stereoscopic image pair, which can be solved by the graph-cut minimization technique [29].

$$E_{total} = \sum_{y=0}^{M-1}\sum_{x=0}^{2(N-1)} (E_{data} + \omega_{sm}E_{sm} + \omega_{co}E_{co}), \quad (16)$$

Image Quality and Convergence Speed Improvement: To generate a high-quality super-resolution stereoscopic image, we first generate a stereoscopic image with the original spatial resolution, therein satisfying the disparity constraint for the target stereoscopic image. Then, we apply our above-defined method (Sec. IV-B and Sec. IV-C) again to generate the high-resolution target stereoscopic image, as shown in Fig. 1.

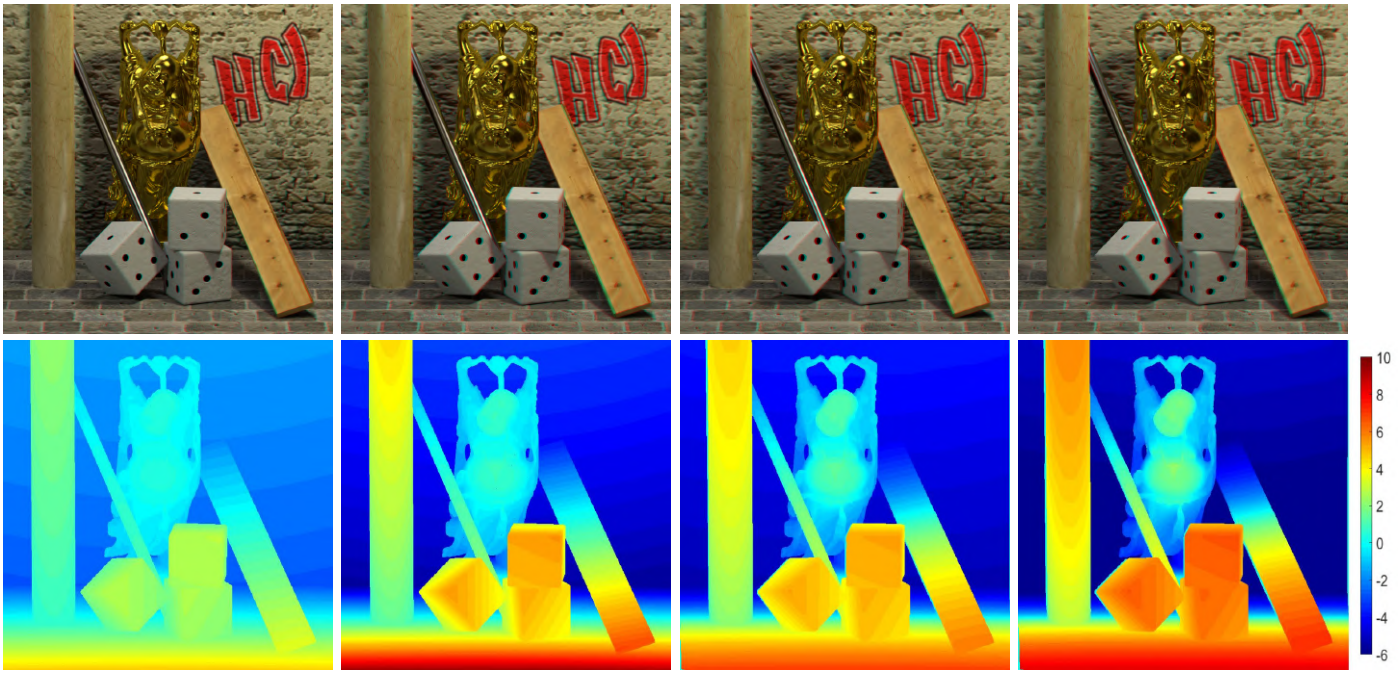
Specifically, when optimizing the high-resolution stereoscopic image, we extend the vectors $V_{p_i}(x',y')$, $\hat{V}_j(x,y)$ and W_d used in the data cost Eq. 13 as follows:

$$V_{p_i}(x',y') = [V_{p_i}(x',y')^T, I^{lw}_{p_i}(x',y')]^T, \quad (17)$$

$$\hat{V}_j(x,y) = [\hat{V}_j(x,y)^T, I^o(x',y')]^T, \quad (18)$$

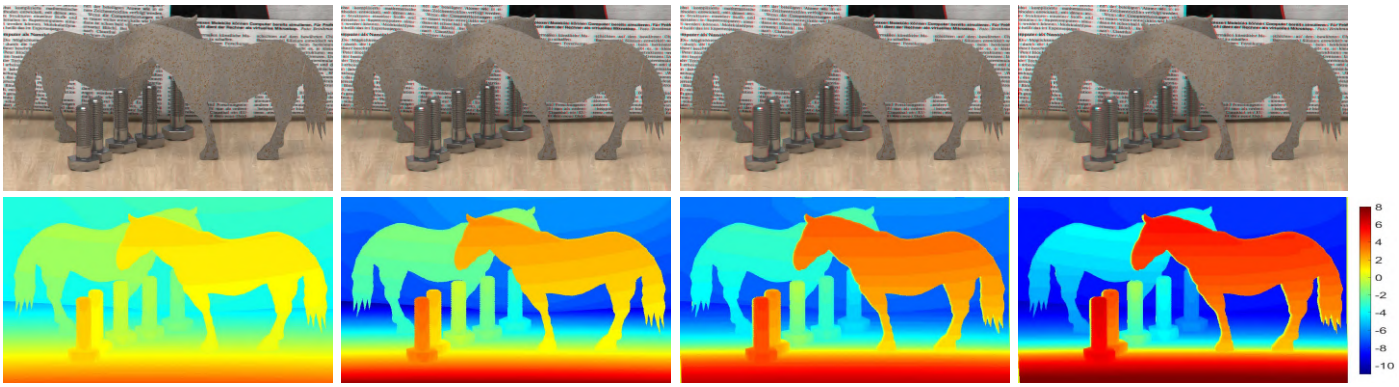
$$W_d(x,y) = [W_d(x,y)^T, \omega_o]^T, \quad (19)$$

where I^o is the optimized stereoscopic image pair with the original resolution and ω_o is a constant weighting parameter.



(a) Reference images (b) Maxmium disparity range (c) Our results (d) Our results

Fig. 6: Nonlinear disparity remapping: (a) shows the stereoscopic image composed of two subaperture views $(4, 2)$ and $(4, 6)$ of the input light field image coming from [41] and the disparity map; (b) shows the stereoscopic image composed of two subaperture views $(4, 0)$ and $(4, 8)$ and the disparity map which is the maximum disparity range of the input light field image; and (c) and (d) are the stereoscopic images and the disparity maps generated by our method at two viewpoints $(4, 2)$ and $(4, 6)$ adopting the nonlinear disparity scaling function (Eq. 5) with parameters ($k_1 = 1$, $k_2 = 4$, $b_1 = 1$) and ($k_1 = 1.5$, $k_2 = 4$, $b_1 = 1$), respectively.



(a) Reference images (b) Maxmium disparity range (c) Our results (d) Our results

Fig. 7: Nonlinear disparity remapping: (a) shows the stereoscopic image composed of two subaperture views $(4, 2)$ and $(4, 6)$ of the light field image coming from [41], and the disparity map; (b) shows the stereoscopic image composed of two subaperture views $(4, 0)$ and $(4, 8)$ and the disparity map which is the maximum disparity range of the input light field image; and (c) and (d) are the stereoscopic images generated by our method and the disparity maps at two fixed viewpoints $(4, 2)$ and $(4, 6)$ utilizing the nonlinear disparity scaling function (Eq. 5) with parameters ($k_1 = 1.25$, $k_2 = 4$, $b_1 = 1$) and ($k_1 = 2$, $k_2 = 4$, $b_1 = 1$), respectively.

We empirically found that by taking such a low-resolution-to-high-resolution pyramid scheme instead of directly computing the high-resolution stereoscopic images, the computational cost of our method is reduced nearly 2-10 times.

V. Experimental Results and Discussion

In this section, we show the experimental results of our method for generating stereoscopic images with given disparity constraints and resolution from light field images. We evaluate our method on the light field data sets from Heidelberg [41] [44] and Stanford [45]. All light field images utilized in our experiments use view indices ranging from $[0, 0]$

to $[8, 8]$.

The parameters in our method are set as $g = 100$, $\omega_{ds} = 1$, $\omega_{dc} = 1$, $\tau_1 = 1.0 \times 10^3$, $\tau_2 = 1.0 \times 10^2$ and $\tau_3 = 1.0 \times 10^2$ in Eq. 12. $\omega_d = 10$, $\omega_l = 20$ in Eq. 13. $\omega_{sc} = 20$, $\omega_{sd} = 40$, $\omega_{sl} = 2.0 \times 10^3$, $T_s = [2.0 \times 10^3, 50, 10, 10]$ in Eq. 14. $\omega_{cc} = 20$, $\omega_{cd} = 4.0 \times 10^2$, $T_c = [2.0 \times 10^3, 50]$ in Eq. 15. $\omega_{sm} = 20$ and $\omega_{co} = 20$ in Eq. 16. $\omega_o = 2$ in Eq. 19. In Eq. 13, for the first iteration, $T_d = [5.0 \times 10^3, 2.0 \times 10^2, 8, 8]$; for the second iteration and onwards, $T_d = [5.0 \times 10^3, 2.0 \times 10^2, 8, 8, 5.0 \times 10^3]$. All our experiments are carried out on a PC with an Intel $i7$ 4GHz CPU and 32GB RAM. We evaluate our method with both linear and nonlinear disparity remappings, at the original resolution and super-resolution. The running time of our method for each stereoscopic view synthesis example ranges from 6 hours to 20 hours, depending on the spatial and angular resolution of input light field images and the resolution of target stereoscopic image.

A. Qualitative Evaluation

In Fig. 5, we show the stereoscopic images generated by our method using the nonlinear disparity scaling function (Eq. 5). From the stereoscopic images and corresponding disparity maps obtained by our method, we can see that comparing with the stereoscopic image directly composed of two subaperture views, our method enlarges the original disparity to enhance the local depth contrast. The depth contrast enhancement can be clearly observed from the local depth changes on the buddha statue in Fig. 5. We show the initial stereoscopic image pair (before our global optimization), in which some black holes, blurring, filling errors highlighted in the red box in the left image and green box in the right image. After our global optimization, there are no obvious distortions or black holes in our final stereoscopic image pair.

In Figs. 6– 7, we show the stereoscopic images generated by our nonlinear disparity scaling function with various parameter settings. From these results, the effectiveness of our method on synthesizing stereoscopic images with flexible global disparity range and locally disparity contrast control can be observed. Users can either globally modify the target disparity range, or locally enhance the disparity contrast by adjusting the parameters of our nonlinear scaling function (Eq. 5). In each example, we show four stereoscopic images and their corresponding disparity maps. In the first column, (a) shows the stereoscopic image composed of subaperture views $(4, 2)$ and $(4, 6)$ and the corresponding disparity map; (b) shows the stereoscopic image composed of subaperture views $(4, 0)$ and $(4, 8)$, and the corresponding disparity maps; (c) and (d) are the stereoscopic images and their corresponding disparity map generated by our method with different parameters for the nonlinear scaling function (Eq. 5). These results demonstrate that our method can flexibly control the disparity range of target stereoscopic image with a nonlinear scaling function (Eq. 5), without introducing any obvious low level artifacts or 3D scene structure distortions. The differences among the colored disparity maps for different stereoscopic images demonstrate the effectiveness of our proposed method.

In Fig. 8, we compare the linear disparity scaling function (Eq. 4) with nonlinear disparity scaling function (Eq. 5) while fixing their target disparity range. The stereoscopic images generated by our method demonstrate that nonlinear disparity scaling function can more effectively enhance depth contrast for target stereoscopic image.

Our method can properly deal with the disocclusion problem in novel view synthesis. Relevant methods [22] and [23] cannot address large black holes in disoccluded regions well. They simple fill such regions with a similar color without synthesizing meaningful textures. Thus, image blurring and filling errors (noise) can always be found in their results [22] [23]. In contrast, our method can properly fill such disoccluded regions with meaningful textures, e.g., Fig. 9. More examples can be found in the supplemental.

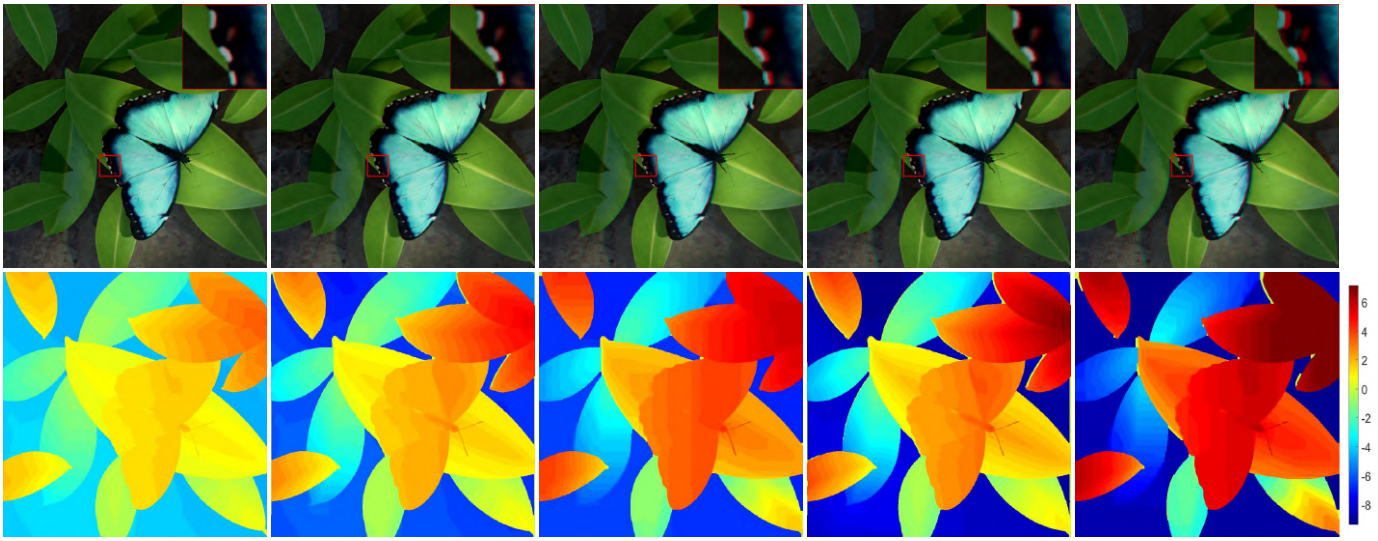
In Fig. 10, the comparison of the stereoscopic images with linear disparity scaling produced by our method using and non-using confidence maps defined in Sec. IV-A3, demonstrate that our method can preserve complex geometry details well when utilizing the proposed confidence maps.

B. Comparison with the State-of-the-Art Method

We show quantitative results of our method for view interpolation and extrapolation in Tabs. I and II, in comparison with the state-of-the-art methods, including [22] [24] [27] [26] [28]. For view interpolation, we utilized the light field view indices ranging from $[0, 0]$ to $[8, 8]$ except subaperture views $(4, 2)$ and $(4, 6)$ to synthesize stereoscopic images at subaperture viewpoints $(4, 2)$ and $(4, 6)$. For view extrapolation, we utilized the light field angular resolution from $[2, 2]$ to $[6, 6]$ to generate stereoscopic images at subaperture viewpoints $(4, 0)$ and $(4, 8)$.

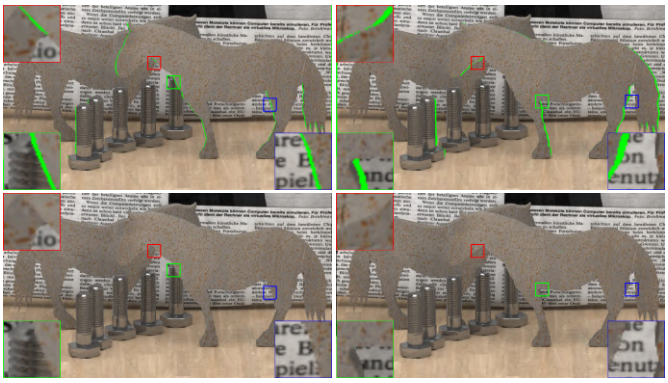
We also show the experimental results of our method comparing with results of Wanner et al. [22] for view interpolation and super-solution in Fig. 12, extrapolation and super-resolution in Fig. 11. In Fig. 12, there is no ground truth disparity map for the light field image truck [45]. Thus, our method takes the disparity maps estimated by [22] as input. It can be observed that in this example there are many noise pixels in results produced by [22], especially in the disocclusion regions near object boundary. In contrast, there are almost no noise in our result. Similar results are shown in Fig. 11, where the experimental results of our method are much better than those produced by [22], without any observable noisy pixels.

In Figs. 13 and 14, we show that our method generates less blurry stereoscopic images from light field images, "Lego Knights" and "Lego Bulldozer", compared with Wanner et al. [22]. As these light field images are captured from real scenes and do not include ground-truth disparity maps, we adopt the state-of-the-art disparity estimation method [46] to obtain disparity maps, which still have much noise and errors. These light field data are captured by large camera array and the baseline and convergence angle between neighboring subaperture views are very large, which makes them very different form other synthetic and real light field images. Therefore, to restore accurate disparity maps and generate high-quality stereoscopic images from these light field images are very challenging. The disparity maps and stereoscopic



(a) Without scaling (b) Linear scaling (c) Nonlinear scaling (d) Linear scaling (e) Nonlinear scaling

Fig. 8: Linear vs. nonlinear disparity scaling. (a) is the stereoscopic image composed of two subaperture views $(4, 2)$ and $(4, 6)$ and the disparity maps. (b) and (c) are results by enlarging the disparity range of (a) 1.5 times with the linear (Eq. 4) and nonlinear (Eq. 5) disparity scaling functions, respectively. (d) and (e) are results by enlarging the disparity range of (a) 2 times with the linear (Eq. 4) and nonlinear (Eq. 5) disparity scaling functions, respectively. By specifying the same disparity range, more obvious depth (contrast) enhancement can be obtained in the results generated by the nonlinear disparity scaling function, (c) and (e), than those produced by the linear disparity scaling function, (b) and (d).



(a) (b)

Fig. 9: Meaningful texture synthesis for disoccluded regions by our method. The first row shows the disoccluded regions in the left and right views marked in red color. (b) shows the optimized results with disoccluded regions filling with meaningful texture by our method.

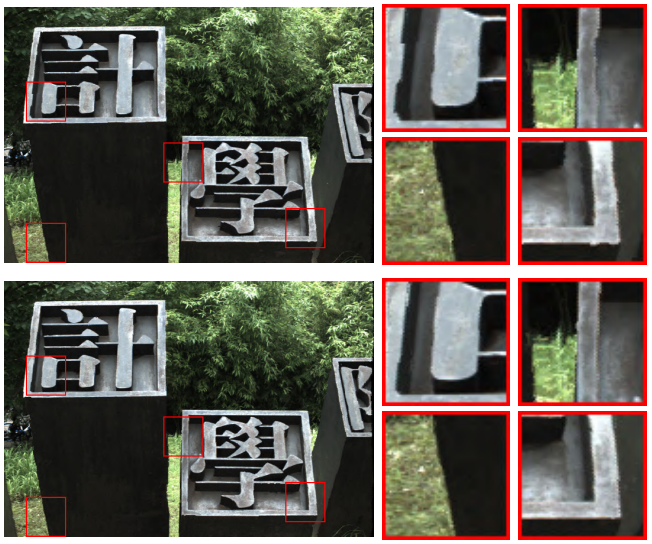
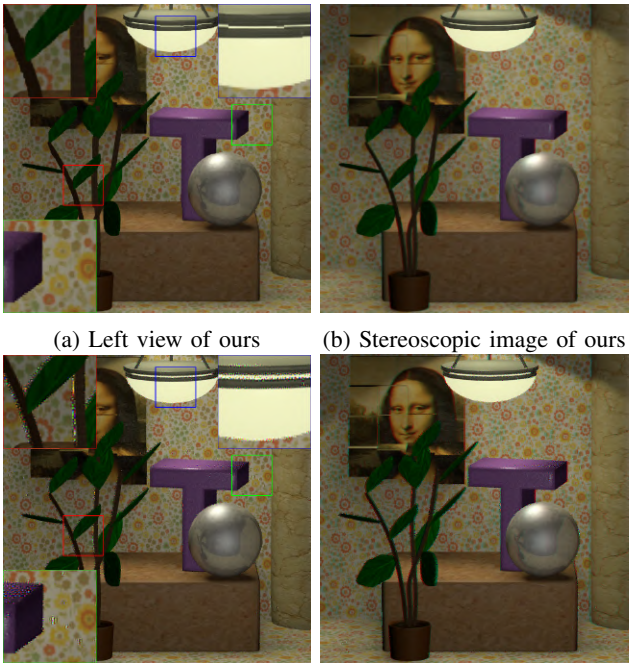


Fig. 10: Stereoscopic image generating from subaperture views $(4, 0)$ and $(4, 8)$ with the linear disparity scaling setting ($k_0 = 10, b_0 = 0$) produced by our method using and non-using confidence maps. From top to bottom, the first row shows the result generated by our method without confidence maps. The second row shows the results generated by our method with the normal confidence maps proposed in our method.

images optimized by our method are more reasonable and visually pleasing than those produced by Wanner et al. [22].

1) Quantitative Evaluation: The main reason for the PSNR values of our method being lower than [22] [27] [28] on view interpolation (Tab. I) is that our method may slightly shift the coordinates of some pixels in the target stereoscopic image. In fact, our method makes a tradeoff between accurate color values and the smooth cost energy term accounting for content preservation (black hole filling and geometry preservation). Our method outperforms the representative deep-learning-based method [26]. In view extrapolation (Tab. II), the PSNR values for results of our method are better than those from the results of [22]. Moreover, the visual quality from human perception of our results is either comparable or better than those of [22], without obvious visual quality decay, e.g., image blurring and noisy pixels.

The PSNR performance of our method is lower than [22] on view interpolation and super-resolution with upsampling



(a) Left view of ours (b) Stereoscopic image of ours

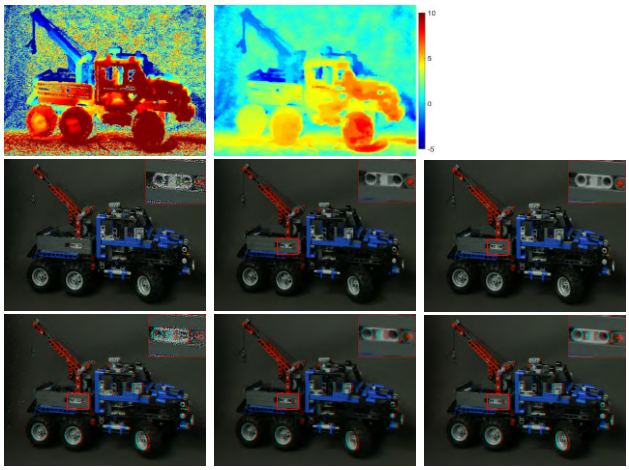
(c) Left view of Wanner [22] (d) Stereoscopic image of [22]

Fig. 11: Comparison between our method and Wanner et al. [22] on view extrapolation and super-resolution. The indices of views ranging from $[2, 2]$ to $[6, 6]$ are utilized to synthesize views $(4, 0)$ and $(4, 8)$ with upsampling factor $2\times$.

factors $2\times$ (Tab. III) and $4\times$ (Tab. V). The PSNR indicator of our method is comparable with the results of [17] on view interpolation and super-resolution with the upsampling factors $2\times$ (Tab. III), but lower than those from the results of [17] with the upsampling factors $4\times$(Tab. V). The PSNR evaluation results on view extrapolation and super-resolution with upsampling factors $2\times$ (Tab. IV) and $4\times$ (Tab. VI) demonstrate that our method outperforms [22]. The following user study evaluation demonstrates that the effectiveness of our method is better than [22] in human perception on view interpolation and extrapolation.

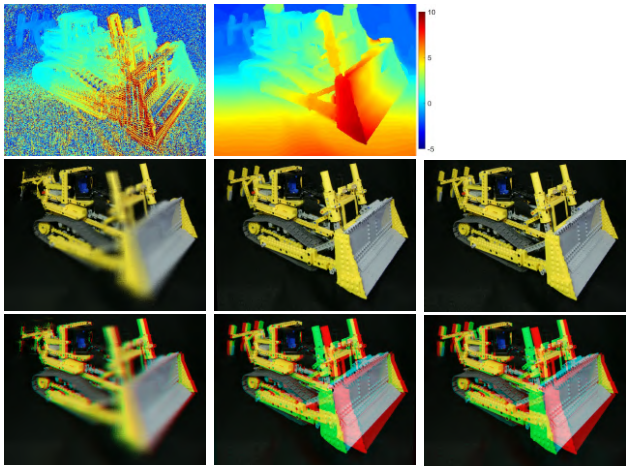
Our method performs much better than [22] when the disparity maps are not accurate enough. Since disparity maps for real light field images are always inaccurate with noise/errors, our method is preferred and effective in most situations, such as Figs. 11–14. Due to the smooth energy term, which can preserve texture structure during texture synthesis, our method performs much better than other methods in view extrapolation with/without super-resolution (Tabs. II, IV and VI), and generates meaningful textures to fill the disoccluded regions (Fig. 9.). Deep-learning-based methods [26] [24] can easily introduce blurring artifacts into the novel synthesized views.

2) User Study Evaluation: We perform a user study by asking 24 volunteers to score the view interpolation and extrapolation results produced by our method and [22], as shown in Tabs. I, II, III and IV (Fig. 15). The volunteers are required to score the images of ground truth (as a reference), our results, and results from Wanner et al. [22]. The results of our method and [22] are shown in random order in the evaluation. Volunteers are asked to score the four metrics



(a) Results of [22] (b) Results of ours (c) GroundTruth

Fig. 12: Comparison between our method and Wanner et al. [22] on view interpolation and super-resolution. The indices of views ranging from $[0, 0]$ to $[8, 8]$, except $(4, 2)$ and $(4, 6)$, are utilized to synthesize views $(4, 2)$ and $(4, 6)$ with upsampling factor $2\times$ and spatial resolution $[640, 480]$. From top to bottom: the first row shows estimated disparity maps for left view; the second row shows optimized target left views; the third row shows target stereoscopic images. The light field image "Lego Truck" is from the light field dataset [45].



(a) Results of [22] (b) Results of ours (c) GroundTruth

Fig. 13: Comparison between our method and Wanner et al. [22] on view extrapolation. The indices of views ranging from $[2, 2]$ to $[6, 6]$ are utilized to synthesize views $(4, 0)$ and $(4, 8)$, with spatial resolution $[768, 576]$. First row shows the optimized disparity maps for target left view. Second row shows the target left views obtained by different methods, and the GroundTruth image. Similarly, third row shows the stereoscopic images. The light field image "Lego Bulldozer" is from the light field dataset [45].

(blurring & noisy, geometric distortion, black hole, perceived image quality) for each image with a value from 0 to 5. For the first three metrics, a smaller value means a better performance, while for the last metric (perceived image quality), a larger value means a better performance.

Approach	Evaluation (PSNR)															
	Buddha	Buddha2	Horse	Medi	Mona	Papi	Still	Maria	Coup	Cube	Table	Town	Truck	Knight	Amethy	Digg
Our Method	39.64	35.32	32.16	32.18	35.34	**37.97**	31.52	24.27	22.25	15.58	30.40	31.31	**36.35**	21.60	23.64	**25.05**
Wanner. [22]	41.49	32.73	35.00	**34.44**	42.50	37.40	**33.65**	31.88	24.93	26.28	35.43	**37.93**	21.55	13.27	22.85	19.64
Wu [27]	**43.84**	**39.94**	**36.48**	33.80	**44.73**	**42.50**	23.14	**42.47**	**33.98**	**35.13**	**38.69**	37.40	31.93	19.55	27.74	19.69
Kalant. [26]	16.67	17.48	18.13	17.43	17.68	19.60	18.26	17.63	19.39	15.51	18.54	18.73	17.21	14.42	19.07	16.14
Vaghar. [28]	41.21	36.64	23.37	32.69	39.68	37.87	21.25	35.88	17.82	18.18	31.02	31.26	30.26	**21.66**	**28.69**	20.02

TABLE I: Evaluation of the light field view synthesis approaches for view interpolation. The test images are from the light field datasets in [41], [44] and [45].

Approach	Evaluation (PSNR)															
	Buddha	Buddha2	Horse	Medi	Mona	Papi	Still	Maria	Coup	Cube	Table	Town	Truck	Knight	Amethy	Digg
Our Method	**38.13**	**34.84**	**32.10**	32.05	**36.75**	**37.97**	**30.66**	**31.53**	**20.96**	17.10	28.50	30.28	**33.61**	**17.57**	**21.63**	**27.15**
Wanner [22]	35.40	26.30	29.30	**32.26**	35.67	30.91	28.13	27.64	17.88	**18.48**	**29.44**	**30.30**	32.62	15.69	20.55	18.53

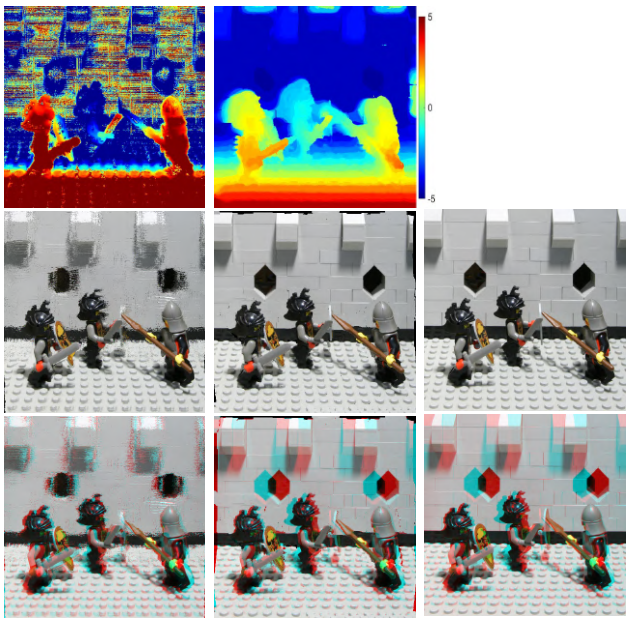
TABLE II: Evaluation of the light field view synthesis approaches for view extrapolation.

Approach	Evaluation (PSNR)															
	Buddha	Buddha2	Horse	Medie	Mona	Papi	Still	Maria	Coup	Cube	Table	Town	Truck	Knight	Amethy	Digg
Our Method	29.47	25.54	23.35	26.89	28.9	29.85	22.29	28.51	**19.24**	15.36	24.36	26.34	**28.18**	18.80	23.98	20.35
Wanner [22]	**37.01**	**31.91**	**28.73**	**31.42**	**34.88**	**34.45**	**26.99**	18.72	17.30	13.68	**27.37**	**30.48**	24.01	**23.62**	24.59	17.52
Yoon [17]	31.07	27.40	21.50	26.35	30.17	31.80	16.66	**29.19**	17.92	**18.20**	25.24	25.11	26.34	20.62	**26.87**	**21.7**

TABLE III: Evaluation of the light field view synthesis approaches for view interpolation and super-resolution with upsampling factor $\times 2$.

Approach	Evaluation (PSNR)															
	Buddha	Buddha2	Horse	Medie	Mona	Papi	Still	Maria	Coup	Cube	Table	Town	Truck	Knight	Amethy	Digg
Our Method	28.68	25.47	23.43	26.82	29.58	29.85	**22.17**	**26.62**	**19.43**	**17.19**	24.32	**25.88**	**29.69**	**17.74**	**21.42**	**23.57**
Wanner [22]	**33.76**	**26.39**	**26.01**	**29.66**	**29.62**	**30.91**	19.26	16.62	15.60	13.23	**26.19**	20.48	24.29	13.63	19.36	17.09

TABLE IV: Evaluation of the light field view synthesis approaches for view extrapolation and super-resolution with upsampling factor $\times 2$.



(a) Results of [22] (b) Results of ours (c) GroundTruth

Fig. 14: Comparison between our method and Wanner et al. [22] on view extrapolation with spatial resolution $[512, 512]$. The description for this example, "Lego Knights", is same to Fig. 13.

The results of our user study evaluation in Fig. 15a demonstrate that our method is comparative with [22] on view interpolation in metrics including blurring, black hole and perceived image quality, although the PSNR values of our method are lower than [22] (Tab. I). Similarly, user study results in Fig. 15b demonstrate that our method outperforms [22] on view extrapolation in blurring & noisy, black hole and perceived image quality, although the PSNR values of both methods are almost the same (Tab. II).

The results of the user study evaluation on view interpolation and super-resolution (Fig. 16a), and view extrapolation and super-resolution (Fig. 16b) also demonstrate that our method outperforms Wanner et.al. [22] in the three metrics (blurring & noisy, black hole and perceived image quality), while slightly higher in the geometric distortion metric.

VI. Conclusion and Future Work

In this paper, we have presented a novel method to generate stereoscopic images from light field images given the disparity scaling constraints and target super-resolution. By applying linear or nonlinear disparity scaling, our method is able to control the global disparity range and adjust the local disparity contrast for target stereoscopic images. Thanks to the rich 3D information recorded by light field images, our method can generate high-quality super-resolution stereoscopic images from arbitrary viewpoints and simultaneously synthesize meaningful textures for the disoccluded regions.

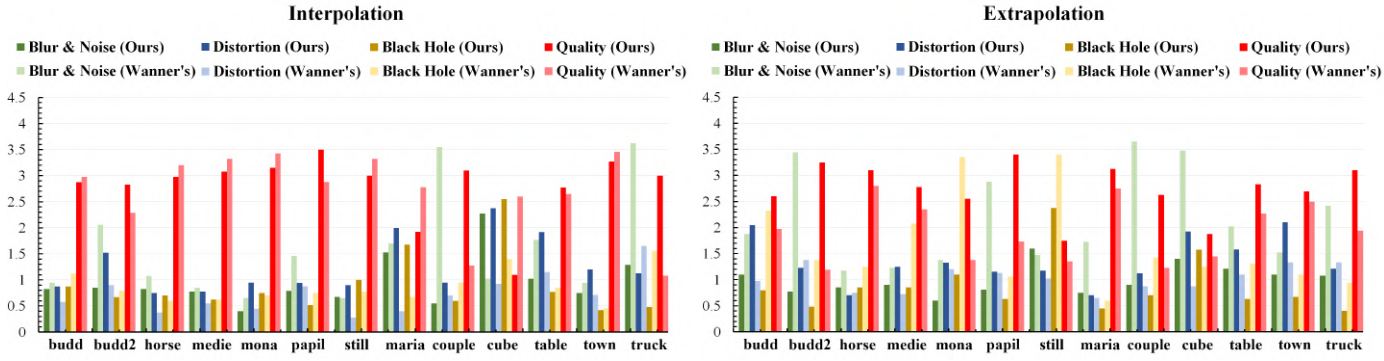
In future work, we would like to improve the computational efficiency of our method to make it run in real time through efficient optimization algorithms and more powerful GPUs. We will also consider improving our method to preserve 3D features, such as large planes, in the output stereoscopic images via sparse 3D reconstruction.

Approach	Evaluation (PSNR)															
	Buddha	Buddha2	Horse	Medi	Mona	Papi	Still	Maria	Coup	Cube	Table	Town	Truck	Knight	Amethy	Digg
Our Method	23.59	25.19	19.49	25.19	22.12	25.62	19.74	23.73	17.02	15.76	20.79	21.57	23.56	15.23	21.74	16.79
Wanner [22]	23.94	**27.71**	**23.00**	**27.71**	28.24	**32.91**	**22.26**	21.07	15.73	13.00	24.33	**26.66**	25.90	17.26	19.78	15.26
Yoon [17]	**30.33**	26.13	22.63	**26.73**	**28.68**	30.96	21.2	**26.29**	**19.33**	**20.15**	**25.28**	25.29	**28.72**	**21.59**	**27.50**	**23.12**

TABLE V: Evaluation of the light field view synthesis approaches for view interpolation and super-resolution with upsampling factor $\times 4$.

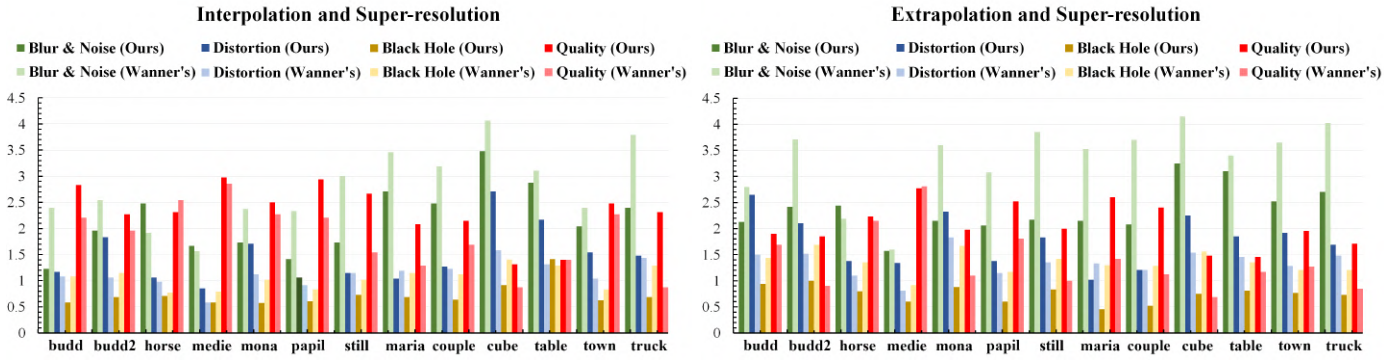
Approach	Evaluation (PSNR)															
	Buddha	Buddha2	Horse	Medi	Mona	Papi	Still	Maria	Coup	Cube	Table	Town	Truck	Knight	Amethy	Digg
Our Method	**26.31**	**19.51**	**19.57**	**24.71**	**21.68**	**25.68**	**18.52**	**22.22**	**16.39**	**15.43**	**19.45**	**21.06**	**24.22**	**16.28**	**21.00**	**18.01**
Wanner [22]	21.99	19.05	16.28	19.26	20.06	22.72	16.19	13.41	14.18	11.83	17.44	17.12	19.85	12.92	16.35	14.73

TABLE VI: Evaluation of the light field view synthesis approaches for view extrapolation and super-resolution with upsampling factor $\times 4$.



(a) View interpolation evaluation (b) View extrapolation evaluation

Fig. 15: User study evaluation on (a) view interpolation and (b) view extrapolation.



(a) View interpolation and super-resolution evaluation (b) View extrapolation and super-resolution evaluation

Fig. 16: User study evaluation on (a) view interpolation and super-resolution ($2\times$), and (b) extrapolation and super-resolution ($2\times$).

REFERENCES

[1] C. Hahne, A. Aggoun, V. Velisavljevic, S. Fiebig, and M. Pesch, "Baseline and triangulation geometry in a standard plenoptic camera," *IJCV*, 2017.

[2] D. Dansereau, O. Pizarro, and S. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *Proc. IEEE CVPR*, 2013, pp. 1027–1034.

[3] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field image processing: An overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 926–954, 2017.

[4] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, "Nonlinear disparity mapping for stereoscopic 3D," *ACM TOG*, vol. 29, no. 3, 2010.

[5] W. J. Tam, F. Speranza, S. Yano, K. Shimono, and H. Ono, "Stereoscopic 3d-tv: Visual comfort," *IEEE Trans. on Broadcasting*, vol. 57, no. 2, pp. 335–346, June 2011.

[6] D. Lanman and D. Luebke, "Near-eye light field displays," *ACM Trans. Graph.*, vol. 32, no. 6, pp. 220:1–220:10, Nov 2013.

[7] S. Du, S. Hu, and R. Martin, "Changing perspective in stereoscopic images," *IEEE TVCG*, vol. 19, no. 8, pp. 1288–1297, 2013.

[8] P. Ndjiki-Nya, M. Koppel, D. Doshkov, H. Lakshman, P. Merkle, K. Muller, and T. Wiegand, "Depth image-based rendering with advanced texture synthesis for 3-d video," *IEEE TMM*, vol. 13, no. 3, pp. 453–465, 2011.

[9] B. Bartczak and et al., "Display-independent 3D-TV production and delivery using the layered depth video format," *IEEE Trans. on Broadcasting*, vol. 57, no. 2, pp. 477–490, 2011.

[10] Y. Zhao, C. Zhu, L. Yu, and M. Tanimoto, *An Overview of 3D-TV System Using Depth-Image-Based Rendering*, 2013, pp. 3–35.

[11] S. Lu, T. Mu, and S. Zhang, "A survey on multiview video synthesis and editing," *Tsinghua Science and Technology*, vol. 21, no. 6, pp. 678–695, 2016.

[12] M. Guttmann, L. Wolf, and D. Cohen-Or, "Semi-automatic stereo

extraction from video footage," in *Proc. IEEE ICCV*, 2009, pp. 136–142.
[13] O. Wang, M. Lang, M. Frei, A. Hornung, A. Smolic, and M. Gross, "Stereobrush: interactive 2D to 3D conversion using discontinuous warps," in *Proc. EG Symp. on Sketch-Based Interfaces and Modeling*, 2011, pp. 47–54.
[14] T. Yan, R. Lau, Y. Xu, and L. Huang, "Depth mapping for stereoscopic videos," *IJCV*, vol. 102, no. 1-3, pp. 293–307, 2013.
[15] "Lytro." [Online]. Available: https://illum.lytro.com/
[16] "Raytrix." [Online]. Available: https://raytrix.de/
[17] Y. Yoon, H. G. Jeon, D. Yoo, J. Y. Lee, and I. S. Kweon, "Learning a deep convolutional network for light-field image super-resolution," in *Proc. IEEE ICCV Workshop*, Dec 2015, pp. 57–65.
[18] M. Rossi and P. Frossard, "Light field super-resolution via graph-based regularization," *CoRR*, vol. abs/1701.02141, 2017.
[19] 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 ICCP*, 2014, pp. 1–10.
[20] M. Alam and B. Gunturk, "Hybrid light field imaging for improved spatial resolution and depth range," *Machine Vision and Applications*, vol. 29, no. 1, pp. 1–12, 2018.
[21] H. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y. Tai, and I. Kweon, "Accurate depth map estimation from a lenslet light field camera," in *Proc. IEEE CVPR*, 2015, pp. 1547–1555.
[22] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE TPAMI*, vol. 36, no. 3, pp. 606–619, March 2014.
[23] Y. Z. Lei Zhang and H. Huang, "Efficient variational light field view synthesis for making stereoscopic 3D images," *Computer Graphics Forum*, vol. 34, no. 7, pp. 183–191, 2015.
[24] Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, and I. S. Kweon, "Light-field image super-resolution using convolutional neural network," *IEEE Signal Processing Letters*, vol. 24, no. 6, pp. 848–852, June 2017.
[25] J. Flynn, I. Neulander, J. Philbin, and N. Snavely, "Deep stereo: Learning to predict new views from the world's imagery," in *Proc. IEEE CVPR*, 2016, pp. 5515–5524.
[26] N. K. Kalantari, T.-C. Wang, and R. Ramamoorthi, "Learning-based view synthesis for light field cameras," *ACM TOG*, vol. 35, no. 6, 2016.
[27] 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 CVPR*, 2017, pp. 1638–1646.
[28] S. Vagharshakyan, R. Bregovic, and A. Gotchev, "Light field reconstruction using shearlet transform," *IEEE TPAMI*, vol. 40, no. 1, pp. 133–147, 2018.
[29] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE TPAMI*, vol. 23, no. 11, pp. 1222–1239, 2001.
[30] V. Kolmogorov and R. Zabin, "What energy functions can be minimized via graph cuts?" *IEEE TPAMI*, vol. 26, no. 2, pp. 147–159, Feb 2004.
[31] C. Kim, A. Hornung, S. Heinzle, W. Matusik, and M. Gross, "Multi-perspective stereoscopy from light fields," *ACM TOG*, vol. 30, no. 6, 2011.
[32] C. Kim, U. Muller, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross, "Memory efficient stereoscopy from light fields," in *Proc. International Conference on 3D Vision*, vol. 1, Dec 2014, pp. 73–80.
[33] T. Yan, S. He, R. W. Lau, and Y. Xu, "Consistent stereo image editing," in *Proc. ACM MM*, 2013, pp. 677–680.
[34] Y. Pritch, E. Kav-Venaki, and S. Peleg, "Shift-map image editing," in *Proc. IEEE ICCV*, 2009, pp. 151–158.
[35] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *IJCV*, vol. 1, no. 1, pp. 7–55, 1987.
[36] S. Wanner and B. Goldluecke, "Globally consistent depth labeling of 4D light fields," in *Proc. IEEE CVPR*, 2012, pp. 41–48.
[37] C. Kim and et.al., "Scene reconstruction from high spatio-angular resolution light fields," *ACM TOG*, vol. 32, no. 4, 2013.
[38] H. Lin, C. Chen, S. B. Kang, and J. Yu, "Depth recovery from light field using focal stack symmetry," in *Proc. IEEE ICCV*, 2015, pp. 3451–3459.
[39] M. Tao, P. Srinivasan, S. Hadap, S. Rusinkiewicz, J. Malik, and R. Ramamoorthi, "Shape estimation from shading, defocus, and correspondence using light-field angular coherence," *IEEE TPAMI*, vol. 39, no. 3, pp. 546–560, 2017.
[40] P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, and H.-P. Seidel, "A perceptual model for disparity," *ACM TOG*, vol. 30, no. 4, 2011.
[41] S. Wanner, S. Meister, and B. Goldluecke, "Datasets and benchmarks for densely sampled 4d light fields," in *Vision, Modeling and Visualization*, 2013, pp. 225–226.
[42] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, "Interactive digital photomontage," *ACM TOG*, vol. 23, no. 3, pp. 294–302, 2004.
[43] V. Kwatra, A. Schödl, I. Essa, G. Turk, and A. Bobick, "Graphcut textures: image and video synthesis using graph cuts," *ACM ToG*, vol. 22, no. 3, pp. 277–286, 2003.
[44] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldlücke, "A dataset and evaluation methodology for depth estimation on 4d light fields," in *Proc. ACCV*, 2016, pp. 19–34.
[45] "The (new) stanford light field archive." [Online]. Available: http://lightfield.stanford.edu/lfs.html
[46] T. Taniai, Y. Matsushita, Y. Sato, and T. Naemura, "Continuous 3d label stereo matching using local expansion moves," *IEEE TPAMI*, vol. 40, no. 11, pp. 2725–2739, 2017.