Generating Stereoscopic Images with Convergence Control Ability from a Light Field Image Pair

Tao Yan, Yiming Mao, Jianming Wang, Wenxi Liu, Xiaohua Qian, Rynson W.H. Lau

***Abstract*—With the advances in commercial light field cameras, light field image processing has attracted considerable attention from researchers. In this paper, we propose a novel method for generating stereoscopic images from a light field image pair with flexible control over the convergence of the virtual stereo cameras. We have developed a light field image-capturing prototype that consists of two horizontally arranged light field cameras (i.e., with their optical axes being parallel to each other). When using our proposed device for image/video capture, stereo photographers can concentrate on how to capture the desired visual experience without being frequently disturbed by having to manipulate the stereo camera parameters, i.e., the convergence angle of a stereo camera. During postprocessing, our method estimates accurate disparity maps for the light field image pair and then generates the target stereoscopic images that satisfy the desired stereo camera convergence requirements by adopting a novel view synthesis method for light field images. We have conducted extensive experiments to demonstrate the effectiveness of our proposed method.**

***Index Terms*—light field image, stereoscopic image, stereo camera convergence, image view synthesis.**

I. Introduction

The light field technique has gained attention from both academia and industry in recent years. Compact commercial light field cameras, e.g., Lytro [1] and RayTrix [2], which have become increasingly popular, use a single 2D photosensor to multiplex the spatial and angular information to form a light field image. Since a light field image can be decoded into a regular array of subaperture views (images) [3], light field images have become one of the most important sources of content for the generation and processing of 3D image/video. There are several applications of light field images, including 3D scene structure inference, image refocusing, segmentation and novel view synthesis [4].

Stereoscopic images/videos are widely used for 3D displays in virtual/augmented reality and robotics. A stereoscopic image consists of a pair of images of the same scene captured from two different viewpoints. However, most stereoscopic images suffer from the accommodation-convergence conflict and a visually uncomfortable disparity range, particularly when they are retargeted for display on various screen sizes. Such problems can be eliminated by disparity scaling [5] [6] or perspective modification methods [7] [8]. Another effective approach to handling these problems is to directly control the interaxial distance of the stereo cameras and the convergence plane for the optical axes of the stereo cameras in the capturing stage [9] [10]. The position of the convergence plane is one of the most important parameters for stereo cameras and determines the positive and negative intervals of the disparity range for the captured stereoscopic images.

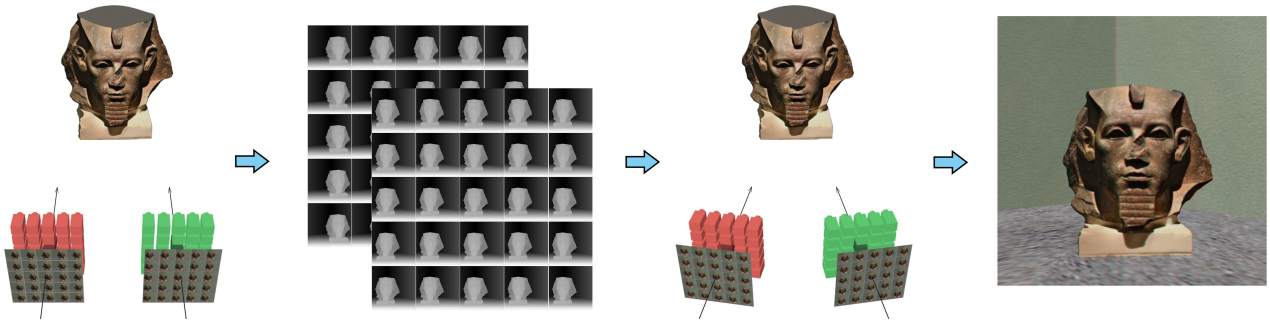
The capture of high-quality stereoscopic images/videos is challenging for photographers [11] [10]. Human stereoscopic experience comes from a combination of several factors, including camera parameters, viewing locations, projector-screen configurations and viewers' psychological factors [10] [5]. In the stereoscopic image/video capture stage, photographers must carefully control the stereo camera baseline and convergence according to the changes in the shots, which requires specific knowledge and experience with stereoscopic movies and the adoption of a specific computational stereo camera system [9]. In this paper, we attempt to replace the difficult control of the convergence angle for stereo camera optical axes with a novel stereoscopic image synthesis method.

Several previous studies have aimed to synthesize stereoscopic images from single light field images [12] [13] [14] by exploiting the redundancy of multiple subaperture views of the same light field image. Due to the narrow baseline of the subaperture views of a light field image, the stereoscopic images that are synthesized from the light field images of a real scene are typically subject to a small disparity range, i.e., small stereoscopic effect. To the best of our knowledge, existing methods for estimating the disparity from a light field image may not always produce very accurate disparity maps. The state-of-the-art methods based on epipolar plane image (EPI) analysis can obtain accurate disparity for the edge pixels of an image, but there is typically substantial noise associated with the pixels in homogeneous regions [15].

In this paper, we propose a novel method to generate stereoscopic images from a light field image pair that can be flexibly controlled for the convergence of a virtual stereo camera pair. An overview of our proposed method is shown in Fig. 1. In the capture stage, we use two horizontally arranged light field cameras with parallel optical axes to capture a pair of light field images. The two central subaperture views of the two light field images can be seen as the original stereoscopic images captured by the stereo cameras with parallel optical

Tao Yan, Jianming Wang and Yiming Mao are with the Jiangsu Key Laboratory of Media Design and Software Technology, Jiangnan University, Jiangsu, China
Wenxi Liu is with the Fuzhou University, Fujian, China
Xiaohua Qian is with the Shanghai Jiao Tong University, Shanghai, China
Rynson W.H. Lau is with the Department of Computer Science, City University of Hong Kong, Hong Kong

This work is supported by the Natural Science Foundation of Jiangsu Province (Grant No. BK20170197), the Fundamental Research Funds for the Central Universities (Grant No. JUSRP1141), and the National Natural Science Foundation of China (Grant No. 61702104).



(a) light field camera pair in parallel (b) estimated disparity maps (c) virtual light field camera pair in toed-in (d) target stereoscopic image

Fig. 1: Overview of our method. Our method uses two horizontally aligned light field cameras to capture a 3D scene to obtain a light field image pair. It then estimates the disparity maps for the light field image pair and generates the target stereoscopic images with the desired convergence, based on novel view synthesis from the perspectively transformed light field image pair.

axes. In this way, stereo photographers can focus on the desired visual experience and avoid being disturbed by having to frequently change the convergence of the stereo cameras according to the shot changes. In the postprocessing stage, our method can estimate accurate disparity maps for the input light field images and manipulate the convergence to synthesize the target stereoscopic images by adjusting the convergence of the virtual stereo cameras via perspective transformation of the light field images. Thus, our method can generate stereoscopic images from light field images with a large disparity range and simultaneously allow users the freedom to flexibly control the disparity range by adjusting the convergence of the virtual stereo cameras.

The main contributions of this paper are as follows:

1) We propose a novel method to generate stereoscopic images with flexible control of the convergence of the virtual stereo cameras from a pair of light field images captured by our proposed capture prototype, i.e., a pair of horizontally aligned light field cameras.
2) We propose a method to estimate accurate disparity maps for a pair of light field images. Our method utilizes the rough disparity maps extracted from both EPIs and corresponding subaperture views of the light field images as prior disparity information.

The rest of this paper is organized as follows. Sec. II reviews previous works relevant to ours. Sec. III presents background information and overviews our method. Sec. IV describes our disparity estimation method for a pair of light field images, and Sec. V introduces our method for synthesizing target stereoscopic images from a pair of light field images. Sec. VI shows experimental results. Finally, Sec. VII concludes our work and discusses possible future research.

II. Related Work

We first review works on stereoscopic image synthesis from light field images. We then discuss works on disparity estimation and novel view synthesis for light field images.

A. Stereoscopic Image Generation from Light Field Images

Recently, some methods [12] [13] [14] based on graph cut optimization [16] and convex variational optimization [17] are proposed to generate stereoscopic images from light field.

Kim et al. [12] proposed a method for producing stereoscopic images with the desired disparity range by adopting a 3D graph cut method on the 3D subspace of EPIs of the light field. This method can process the whole 4D spatiotemporal light field volume to generate the desired stereoscopic video. 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, this method supports only depth adjustments for the selected objects instead of global adjustments.

Subsequently, Kim et al. [13] proposed another method that adopted multiperspective imaging for the synthesis of a stereoscopic image from a light field image. By utilizing a light field image and a manually drawn intended disparity map as the input, this method synthesizes the target view by selecting light rays that fulfill the given disparity constraints. Specifically, this method takes a subaperture view as one view of the target stereoscopic image; the unknown second view is generated by solving a convex variation optimization problem to determine the subaperture view that each pixel should be taken from. Therefore, the disparity range of the target stereoscopic image is constrained by the disparity range of the input light field image.

Zhang et al. [14] proposed a method to synthesize stereoscopic images from an input light field image with the desired disparity constraint. By specifying one subaperture view of the input light field image, this method adopts variational optimization to synthesize another view of the target stereoscopic image by adopting a weighted view interpolation or extrapolation. The authors adopted linear, nonlinear, and artistic disparity scaling on the original disparity range. If the viewpoint of the new synthesized image is beyond the range of the light field view, this method allows the disparity range of the target stereoscopic image to be larger than that

of the input light field image. Moreover, an efficent solution for optimizing the convex total variation is proposed to speed up color computation of every pixel in the target view.

All these methods can generate a desired stereoscopic image from a single light field image. However, the disparity range for the target stereoscopic image is always constrained by the narrow disparity range of the input light field image. Additionally, the quality of the generated stereoscopic image relies heavily on the accuracy of the estimated disparity maps for the input light field images.

B. Disparity Map Estimation for Light Field Images

Various methods have been proposed to estimate depth from light field images. These methods can be mainly divided into two categories: optimization-based methods and deep-learning-based methods.

Optimization-based methods: Wanner et al. [15] proposed an efficient disparity map estimation method that included three steps: local depth labeling in the EPI space that utilized the structure tensor technique, consistent EPI depth labeling, and depth integration from the horizontal and vertical EPI slices. The structure tensor technique can estimate rough disparity maps by computing the directions of the slopes in EPIs. Although the disparity of pixels on edges is accurate, other pixels in homogeneous regions may be assigned incorrect disparity values that are difficult to rectify in the subsequent total-variation-based optimization procedure, i.e., consistent EPI depth labeling and global integration.

Lin et al. [18] proposed a method to recover the depth map from a light field image by exploiting two features of the light field focal stack. One feature is the property that nonoccluding pixels exhibit symmetry along the focal depth dimension centered over the in-focus slice. The other is the data consistency metric of a Markov random field (MRF), which is used to measure the difference between the focal stack estimated from the hypothesized disparity map and the full-focus central image and the focal stack synthetically generated from the input light field image.

Jeon et al. [19] adopted multiview stereo-based matching with a phase-based subpixel shift for the estimation of light field image disparity. Sajjadi et al. [20] proposed a generative model for a light field image, which was fully parameterized by the central subaperture view and its corresponding disparity map. The disparity maps recovered by this method are more accurate than those recovered by other methods, particularly for the light field images of a real scene captured by a compact commercial light field camera.

All these methods estimate the disparity maps from a single light field image based on multi-label optimization techniques. However, substantial noise/errors remain in the estimated disparity maps.

Deep-learning-based methods: Hazirbas et al. [21] proposed an autoencoder-style convolutional neural network to estimate the disparity map from a focal stack of a light field image of real scene. To train the proposed deep learning network, the authors built a dataset that included a large number of light field images and the corresponding registered ground truth depth maps recorded with an RGB-D sensor.

Heber et al. [22] presented a novel U-shaped autoencoder-style deep learning network to extract geometric information from light field data. This network takes the 3D subsets of the 4D light field, i.e., 3D EPI volumes, as input data. This network then uses 3D convolutional layers to propagate information from two spatial dimensions and one directional dimension of the light field. This method can reduce depth artifacts and maintain clear depth discontinuities.

Guo et al. [23] proposed a unified learning-based technique that simultaneously utilizes binocular stereo cues and monocular focus cues for depth inference. This network adopts a pair of focal stacks as the input to emulate human perception. The authors constructed three individual networks: a FocusNet to extract depth from a single focal stack, an EDoFNet to obtain the extended depth of the field image from the focal stack, and a StereoNet to conduct stereo matching. These three networks are integrated into a unified solution to obtain the final depth maps. The EDoF image from EDoFNet serves to both guide the refinement of the depth from FocusNet and provide inputs for StereoNet.

Since deep-learning-based disparity estimation methods are typically trained on light field images of a limited number of real or virtual scenes, they always produce disparity maps that are insufficiently accurate and cannot outperform the optimization-based methods when processing light field images of a new scene. Our method can estimate high-accuracy disparity maps from a light field image pair based on the multi-label optimization technique [16] for MRF optimization.

C. View Synthesis from Light Field Images

The existing methods for the inter/extrapolation of views in light field images can be classified into two categories: total-variation-based methods [15] and deep convolutional neural network (DCNN) based methods [24] [25] [26] [27]. Because of rich 3D information recorded in light field images, these methods taking light field images as input can always produce higher-quality novel views than traditional new view synthesis (navigation) methods [28] [29] based on image-based rendering(IBR) and a set of 2D images.

Wanner et al. [15] proposed a framework for light field analysis based on the total variation. This method can estimate disparity maps and synthesize super-resolution novel views from a light field image. The authors proposed a variational model to synthesize the super-resolved novel views. This method, which works at the subpixel level and is very efficient, is a state-of-the-art method for synthesizing super-resolution novel views in light field images. However, the method does not work well for view extrapolation and the disocclusion problem, due to its insufficient constraints for the total variation, and it it also easily affected by inaccurate disparity maps for the input light field image.

Recently, representative works based on DCNN [24] [25] demonstrated the synthesis of a novel view image with a super-resolution effect in both the spatial and angle domains. One method [24] upsamples the spatial and angular resolution of a light field image. First, the spatial resolution of each subaperture view is increased and the local details of the

image content are enhanced by adopting a spatial super-resolution network with four layers. Then, novel views are generated between subaperture views by adopting an angular super-resolution network with four layers. These two networks are trained independently and then fine-tuned via end-to-end training. Another method [25], which is an extension of the previous method [24], improves the efficiency of the training procedure and the quality of the angular super-resolution results by adopting weight sharing in the angular super-resolution network. However, these two works use only the two adjacent subaperture views for angular super-resolution, which means that they underexploit the information provided by the light field image. In addition, this general approach achieves a fixed super-resolution rate of only $\times 2$.

Flynn et al. [26] designed a DCNN that consists of a selection tower and a color tower to generate novel views for directly synthesizing stereoscopic images from multiview input. The selection tower predicts an approximate depth for each pixel in the output image and determines the source pixels from the input images that can be used as candidate pixels to generate pixels for the output image. The color tower uses all relevant source pixels to predict a color for that output pixel. This deep architecture is trained from a large number of posed image sets (i.e., images with known camera parameters), such as Google's Street View image database and the KITTI dataset. Using this method, pixels from neighboring views of a scene are presented to the network, which are used to produce the pixels of the novel view. This method can be extended to synthesize novel views from light field.

Kalantari et al. proposed a representative work for novel view synthesis in light field images [27] that primarily involves two steps, each implemented with one simple four-layers DCNN model. First, this method predicts the target disparity map for a specified novel view with one DCNN model. Second, the prewarped subaperture views and the target disparity map from the first step are taken as inputs to synthesize the target view with another DCNN model.

Since novel view synthesis methods based on deep learning always rely on an accurate disparity map and pretraining on similar scenes, we adopt the total-variation-based novel view synthesis method [15] for our new view synthesis in this work.

III. Background and Overview of Our Method

Interaxial distance and *convergence* are the two most important parameters of a stereo camera. The *interaxial distance* of a stereo camera pair determines the global disparity range of produced stereoscopic images. The *convergence* of a stereo camera pair determines the position of the focal plane for the shooting scenes, i.e., it determines the positive and negative disparity distribution for a stereoscopic image. Our proposed method focuses on controlling the convergence of a virtual stereo camera pair by synthesizing the desired stereoscopic images from a light field image pair.

Generally speaking, a large baseline means a large disparity range for the captured stereoscopic images. However, the disparity range of a stereoscopic image should be within the comfortable disparity range for a specific stereo viewing condition. The depth of a pixel perceived by the viewer is determined by its disparity, the physical resolution of target stereo display, the distance from the display to the viewer and the interpupillary distance of the viewer [5] [30] [31].

The light field is becoming a very active research topic in the computer vision community. A light field camera can record a regular array of multiperspective images of the same scene in a single photographic exposure. Two neighboring subaperture views in the horizontal or vertical direction usually have equal baselines and focal lengths among other parameters. However, the baseline is typically very small, especially for a commercial light field camera. All subaperture views of a light field image have optical axes that converge at a convergence point, where the optical axis of the light field camera intersects with the focal plane. Therefore, the light field camera enables the points on the focal plane to have zero disparity values in the estimated disparity maps.

The limitation of a low disparity range always occurs when a stereoscopic images is generated from a single light field image. On the other hand, although accurate disparity values for pixels at the edges of images can be easily obtained by adopting the structure tensor technique [15], the disparity maps estimated by the state-of-the-art methods [15] [19] are not sufficiently accurate, especially for pixels in homogeneous regions of EPIs of light field images.

In this work, we propose a method to generate stereoscopic images from a light field image pair, allowing the user to flexibly control the convergence of the virtual stereo camera pair for the target stereoscopic images. Our method has two main steps as follows.

The first step is to estimate the disparity maps for a light field image pair. A unified optimization framework is designed to jointly optimize the disparity maps for a light field image pair. Two types of initial disparity maps separately extracted from EPIs and the corresponding subaperture views of the input light field image pair are taken as prior disparity information and integrated into this optimization framework. We treat these two types of disparity maps as complementary information for the edges and homogeneous regions of the image content.

The second step is to generate the desired stereoscopic images that satisfies the convergence requirements for the virtual stereo camera pair from a light field image pair based on the previously optimized disparity maps. After the convergence of the target virtual stereo camera pair has been specified, our method computes the stereo viewpoint pair for the target stereoscopic image pair and performs the corresponding perspective transformations for the input light field image pair. The target stereoscopic image pair is then generated based on a state-of-the-art novel view synthesis method [15] by taking all subaperture views of the warped light field images as inputs.

IV. Disparity Map Estimation

We design a unified optimization framework to estimate the disparity maps for a light field image pair. By taking advantage of these two types of disparity maps, which are separately estimated from EPIs, and the corresponding subaperture views

of the input light field image pair, our method can produce accurate disparity maps. The benefits accrue from two aspects. On one hand, this method uses the disparity maps estimated from the corresponding subaperture views to refine/correct the noise in the disparity map estimated from the EPIs. On the other hand, this method optimizes the disparity maps for a light field image pair simultaneously and enforces the disparity coherence between different subaperture views of inter- and intralight field images.

Our disparity estimation method involves three steps. First, we preprocess the input light field images to obtain the initial coarse disparity maps from EPIs [15], the disparity maps inferred from the corresponding subaperture views of a light field image pair [32], and the gradient maps for all subaperture views. Second, we estimate the intrinsic camera parameters of the input light field image pair to register these two types of disparity maps. Finally, our proposed optimization framework is used to produce accurate disparity maps for the input light field image pair.

A. Light Field Image Preprocessing

We denote the input light field image pair as L_1 and L_2. The preprocessing for our disparity estimation method can be described as follows.

First, we extract the raw disparity maps $\bar{D}_1$ and $\bar{D}_2$ from the EPIs of L_1 and L_2 by adopting the structure tensor technique [15]. The confidence maps corresponding to the these disparity maps are set as C_1 and C_2. Since stereo matching is an extensively studied topic in computer vision community, there are a lot of excellent methods which can produce high-quality disparity maps for stereo images [33] [32]. We estimate the disparity maps $\hat{D}_1(u,v)$ and $\hat{D}_2(u,v)$ for the respective subaperture views $L_1(u,v)$ and $L_2(u,v)$ by adopting the state-of-the-art stereo matching method [32].

Second, we compute the gradient maps for the light field images L_1 and L_2, which are set to G_1 and G_2, respectively. Then, we normalize G_1 and G_2 to produce the saliency maps for L_1 and L_2, which are denoted as S_1 and S_2, respectively.

Third, we employ a cross-check for $\hat{D}_1$ and $\hat{D}_2$. We keep the consistent disparity values and remove the inconsistent ones from $\hat{D}_1$ and $\hat{D}_2$. Two sets of mask maps, M_1 and M_2, are used to indicate the remaining consistent disparity values in $\hat{D}_1$ and $\hat{D}_2$. Similarly, we conduct a cross-check for disparity maps $\bar{D}_1$ and $\bar{D}_2$ and utilize two sets of mask maps, P_1 and P_2, to indicate the remaining consistent disparity values.

B. Light Field Image Registration

We solve for three key parameters of the input light field image pair. B is the baseline between two central subaperture views, and b represents the baseline between two neighboring subaperture views of the same light field image in the horizontal or vertical direction. Another key parameter is d_s, which represents the shift value of the projected subaperture view on the photosensor of the light field camera. Since the standard light field camera [34] and a simulated light field camera with blender [35] always adopt a tilted stereo camera model for any two selected subaperture views [34], d_s is important parameter for analyzing the disparity of light field images.

The relationship between the disparity and depth of a pixel in a light field image can be determined as follows [34]:

$$d = \frac{bf}{z} - d_s, \tag{1}$$

where d is the disparity, z is the depth, and f is the focal length of the subaperture views of a light field image.

We select the two central subaperture views of L_1 and L_2 to compute our light field image registration, which we designate as $L_1(u_c,v_c)$ and $L_2(u_c,v_c)$. Since the rough disparity maps $\bar{D}_1(u_c,v_c)$ and $\bar{D}_2(u_c,v_c)$ contain substantial noise, and the disparity inferred from EPIs on strong slopes is more accurate than that from any other pixels in homogeneous regions, we compute the two edge maps $E_1(u_c,v_c)$ and $E_2(u_c,v_c)$ for $\bar{D}_1(u_c,v_c)$ and $\bar{D}_2(u_c,v_c)$ by simply thresholding the corresponding gradient maps $G_1(u_c,v_c)$ and $G_2(u_c,v_c)$.

The relationship of $\bar{D}_1$ and $\hat{D}_1$ for L_1 can be modeled as:

$$\bar{D}_1(u,v,x,y) = \mu_0 \hat{D}_1(u,v,x,y) - \mu_1, \tag{2}$$

where $\mu = [\mu_0, \mu_1]$, $\mu_0 = b/B$ and $\mu_1 = d_s$. The relationship between $\bar{D}_2$ and $\hat{D}_2$ is similarly acquired.

The optimal value for the vector variable μ can be obtained by minimizing the following objective function:

$$\begin{aligned}\min \sum_{(x,y)} & m_1(x,y)\|[\hat{D}_1(u_c,v_c,x,y),-1]\mu^T - \bar{D}_1(u_c,v_c,x,y)\|_2^2 \\ & + m_2(x,y)\|[\hat{D}_2(u_c,v_c,x,y),-1]\mu^T - \bar{D}_2(u_c,v_c,x,y)\|_2^2,\end{aligned} \tag{3}$$

where

$$m_1(x,y) = E_1(u_c,v_c,x,y)M_1(u_c,v_c,x,y)P_1(u_c,v_c,x,y), \tag{4}$$

$$m_2(x,y) = E_2(u_c,v_c,x,y)M_2(u_c,v_c,x,y)P_2(u_c,v_c,x,y), \tag{5}$$

and (x,y) represents a pixel on the strong edge of the central subaperture view of L_1 and L_2 with a high-confidence disparity value and simultaneously satisfies the cross-check for the two types of disparity maps.

To effectively solve the objective function in Eq. 3, we rewrite it as follows:

$$\begin{aligned}&\min \|Au - Y\|_2^2 \\ &= (Au-Y)^T(Au-Y) \\ &= u^T A^T A u - 2Y^T A u + Y^T Y,\end{aligned} \tag{6}$$

where $u_0 > 0$ and $u_1 \geq 0$ are hard constraints for the objective function, A is a matrix that consists of $[\hat{D}_1(u_c,v_c,x,y),-1]$ and $[\hat{D}_2(u_c,v_c,x,y),-1]$ as defined in Eq. 3, and Y is a vector that consists of $\bar{D}_1(u_c,v_c,x,y)$ and $\bar{D}_2(u_c,v_c,x,y)$ as defined in Eq. 3. We solve the objective function as a standard convex optimization problem to find the optimal value for u.

Once the optimal values for b/B and d_s have been obtained, we can successfully build the relationship between the two types of disparity maps as defined in Eq. 2. Then, the disparity

maps $\hat{D}_1$ and $\hat{D}_2$ can be mapped to the disparity range for the input light field images as follows:

$$\begin{aligned}\hat{D}_1^t &= \frac{b}{B}\hat{D}_1 - d_s,\\ \hat{D}_2^t &= \frac{b}{B}\hat{D}_2 - d_s.\end{aligned} \tag{7}$$

C. Disparity Map Estimation Based on MRF

We propose an optimization framework to estimate the disparity maps for the input light field image pair. In this way, our method is able to produce accurate and coherent disparity maps. Our optimization framework is designed based on MRF and solved by the graph cut algorithm [16]. The related energy terms are defined as follows.

1) Data Cost: Our optimization framework takes the two types of disparity maps as prior information. The data cost term for pixel $p = (u, v, x, y)$ assigned label l_p is defined as:

$$E_d(p, l_p) = E_{epi}(p, l_p) + E_{ste}(p, l_p), \tag{8}$$

where

$$E_{epi}(p, l_p) = w_1 \min(\|f(\bar{D}(p)) - l_p\|_2^2, \tau_1), \tag{9}$$

$$E_{ste}(p, l_p) = w_2 \min(\|f(\hat{D}^t(p)) - l_p\|_2^2, \tau_1). \tag{10}$$

Here, $f(d)$ is adopted to map the continuous disparity range to a discrete label space, which is defined as:

$$f(d) = \frac{d - d_{min}}{l_u}. \tag{11}$$

where l_u is a constant parameter, d_{min} is the minimum disparity value, and w_1 and w_2 are weighting parameters defined as:

$$\begin{aligned}w_1(p) =& K_1 * S(p) * C(p) * R(p),\\ w_2(p) =& K_2 * M^{'}(p) * R(p),\end{aligned} \tag{12}$$

where K_1 and K_2 are constant weighting parameters. The function R is used to measure the coherence between two types of disparity maps $\bar{D}$ and $\hat{D}^t$, and defined as:

$$R(p) = \max(\exp(-|\hat{D}^t(p) - \bar{D}(p)|), \tau_\gamma). \tag{13}$$

$M^{'}$ is used to measure the confidence of disparity values in $\hat{D}_1$ and $\hat{D}_2$, which is defined as:

$$M^{'}(p) = \begin{cases} 1.0 & \text{if } M(p) == 1 \\ \tau_\alpha & \text{else} \end{cases}, \tag{14}$$

where τ_1, τ_r and τ_α are constant parameters of the truncation functions.

2) Smooth Cost: To enforce the smoothness of the disparity changes between neighboring pixels, $p = (u, v, x, y)$ and $q = (u, v, x_n, y_n)$, in the same subaperture view, we define the smooth energy term as:

$$E_s(p, q, l_p, l_q) = \sum_{q \in N_p} W_s^T |V(q) - V(p)| * \min(|l_q - l_p|, \tau_2), \tag{15}$$

where $V = [I, G, \bar{D}]$, $W_s = K_3[w_{sc}, w_{sg}, w_{sd}]$ is a constant weighting vector, and τ_2 is a constant threshold parameter.

3) Coherence Cost for Subaperture Views of the Same Light Field Image: We define an energy term to enforce the disparity coherence for neighboring subaperture views of the same light field image. If a pixel $p = (u, v, x, y)$ and another pixel q ($q = (u, v + 1, x, y - \bar{D}(u, v, x, y))$ or $q = (u+1, v, x-\bar{D}(u, v, x, y), y)$) are in two separate neighboring subaperture views and satisfy the following relationship, we add a nondirect edge between them to represent the disparity coherence constraint for these two pixels:

$$|\bar{D}(p) - \bar{D}(q)| < \tau_3, \tag{16}$$

where τ_3 is a constant parameter.

The coherence energy cost for the two pixels p and q taking the labels l_p and l_q is defined as

$$E_{ci}(p, q, l_p, l_q) = \begin{cases} K_4 & \text{if } l_p == l_q \\ 0 & \text{else} \end{cases}. \tag{17}$$

4) Coherence Cost for Corresponding Subaperture Views of the Light Field Image Pair: An energy term is defined to enforce the disparity coherence for the corresponding subaperture views of L_1 and L_2. We denote a pixel $p = (u, v, x, y)$ in L_1 and another pixel $q = (u, v, x - \hat{D}_1(u, v, x, y), y)$ in L_2. If these pixels satisfy the following relationship, we conclude that they are a matched pixel pair in the corresponding subaperture views.

$$|\hat{D}_1(p) - \hat{D}_2(q)| < \tau_4, \tag{18}$$

where τ_4 is a constant parameter.

The coherence energy cost for the two matched pixels p and q separately within the two corresponding subaperture views of L_1 and L_2 is defined as

$$E_{ce}(p, q, l_p, l_q) = \begin{cases} K_5 & \text{if } l_p == l_q \\ 0 & \text{else} \end{cases}. \tag{19}$$

Total Energy Function: Finally, we minimize the following objective function to obtain the optimal disparity maps:

$$(D_1, D_2) = \arg\min E_d + E_s + E_{ci} + E_{ce}, \tag{20}$$

where D_1 and D_2 are the output disparity maps estimated by our method.

V. Stereoscopic Image Synthesis

Based on the input light field image pair and its disparity maps estimated in Sec. IV, our method can generate target stereoscopic images with the desired convergence requirement.

1) Convergence Requirement for a Virtual Stereo Camera Pair: We estimate the focal distance for the input light field image pair by adopting the following function:

$$z_0 = \frac{bf}{d_s}, \tag{21}$$

where z_0 is the focal distance and b, f and d_s are parameters of the input light field image pair, as discussed in Sec. IV-B.

Referring to the original focal distance as a reference, the focal distance for the target stereoscopic image can be flexibly adjusted by the user. In addition, the position of the convergence point on the focal plane can be shifted flexibly.

Once the convergence point, $P_c = (X_c, Y_c, Z_c)$, for the virtual stereo camera/viewpoint pair of the target stereoscopic image has been specified, the rotation angles for the target stereo viewpoint pair can be easily obtained. We refer to these angles for the left and right viewpoints of the target stereo viewpoint pair as β_1 and β_2, and compute them as:

$$\begin{aligned} \beta_1 &= \arctan \frac{Z_c}{X_c - X_{o1}}, \\ \beta_2 &= \arctan \frac{Z_c}{X_c - X_{o2}}, \end{aligned} \tag{22}$$

where $O_1 = (X_{o1}, Y_{o1}, Z_{o1})$ and $O_2 = (X_{o2}, Y_{o2}, Z_{o2})$ are the optical centers of the central subaperture view of light field images L_1 and L_2, $Y_c = Y_{o1} = Y_{o2} = 0$.

2) Stereoscopic Image Synthesis: To generate the target stereoscopic image pair that satisfies the desired convergence requirement, our method first conducts a perspective transformation of the light field images L_1 and L_2. Then, the target stereoscopic image pair is generated based on the perspective transformed light field image pair.

Based on the optimized disparity maps and selected key parameters for the input light field image pair, our method can generate the perspective transformed light field images. The depth of a pixel $p = L(u, v, x, y)$ can be determined by its disparity $d = D(u, v, x, y)$, as defined in Eq. 1.

The viewpoint for the subaperture view (u, v) of the input light field can be defined as:

$$\begin{cases} X_o(u,v) = (u - u_c)b \\ Y_o(u,v) = -(v - v_c)b \\ Z_o(u,v) = 0 \end{cases}, \tag{23}$$

where $(u_c, v_c, 0)$ is the viewpoint for the central subaperture view of an input light field image.

Then, after the perspective transformation of the light field image, the viewpoint for the subaperture view (u, v) of the transformed light field is defined as:

$$\begin{cases} X_t(u,v) = (u - u_c)b\cos\theta \\ Y_t(u,v) = -(v - v_c)b \\ Z_t(u,v) = (u - u_c)b\sin\theta \end{cases}, \tag{24}$$

where θ is the rotation angle of the viewpoint plane of the light field around the Y axis. To obtain the desired convergence control, θ is equal to β_1 for the left view, and θ is equal to β_2 for the right view of our target stereoscopic image, as shown in Eq. 22.

Our method performs perspective transformation for all subaperture views of the light field images based on the depth-image-based rendering (DIBR) concept. First, the method projects every pixel in a subaperture view to the 3D space. Then, it projects a point in the 3D scene back to the corresponding subaperture view of the transformed light field image. In this way, we generate two sets of warped subaperture views for the transformed light field image pair, which are marked as L_1^t and L_2^t, while the initially warped views may contain some small black holes that must be inpainted through a further optimization step.

After all subaperture views of L_1 and L_2 have been transformed to the corresponding subaperture views, our method synthesizes the target stereoscopic image whose virtual stereo viewpoint pair is defined at the two central subaperture viewpoints of the light field images pair that was transformed above. We adopt the convex optimization framework for novel view synthesis in the light field image [15] to generate our target stereoscopic image pair. This method is adopted to produce super-resolved novel views without black holes from the warped subaperture views of L_1^t and L_2^t. This method can produce new visually pleasing views for our stereoscopic image generation.

VI. Experimental Results and Discussion

To evaluate the effectiveness of our method, we conduct experiments on a set of light field image pairs from real scenes and a set of synthetic light field image pairs with the corresponding ground truth disparity maps and desired stereoscopic images.

The parameter settings for our disparity estimation (Sec. IV-C) are listed in Table I. Every light field image pair from a real scene is captured by a Lytro Illum camera that shifts on a horizontal shelf without rotation. Each light field image pair from a virtual scene with the ground truth disparity maps and stereoscopic image pairs is rendered with the blender software [35]. light field image pair has a baseline B between the two central subaperture views. For each light field image pair, we first estimate its disparity maps as discussed in Sec. IV. We then generate the desired stereoscopic images according to the convergence requirements for the virtual stereo viewpoint pair, as discussed in Sec. V.

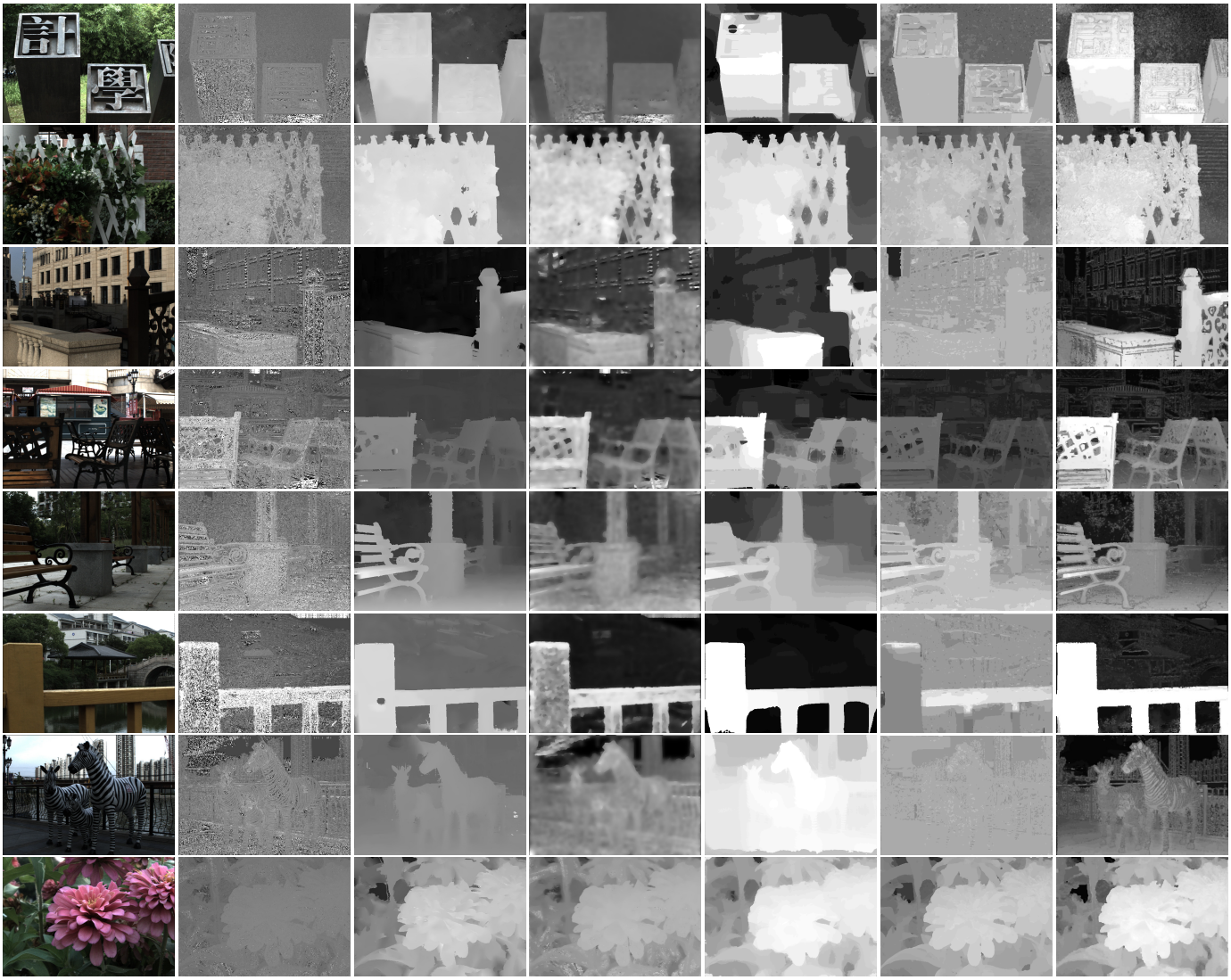
Eq. #	Parameters Setting
Eq. 9	$\tau_1 = 6^2$
Eq. 10	$\tau_2 = 6^2$
Eq. 11	$l_u = 0.02$
Eq. 12	$K_1 = 10^3$ for real scenes or $K_1 = 2 \times 10^2$ for virtual scenes $K_2 = 10^2$
Eq. 13	$\tau_\gamma = 0.02$
Eq. 14	$\tau_\alpha = 0.02$
Eq. 15	$K_3 = 10$, $\tau_2 = 6$ $[w_{sc}, w_{sg}, w_{sd}] = [0.3, 0.3, 0.4]$
Eq. 16	$\tau_3 = 0.03$
Eq. 17	$K_4 = 10^2$
Eq. 18	$\tau_4 = 0.6$
Eq. 19	$K_5 = 10^2$

TABLE I: Parameter settings of our disparity estimation and stereoscopic image generation method.

A. Disparity Map Estimation Results

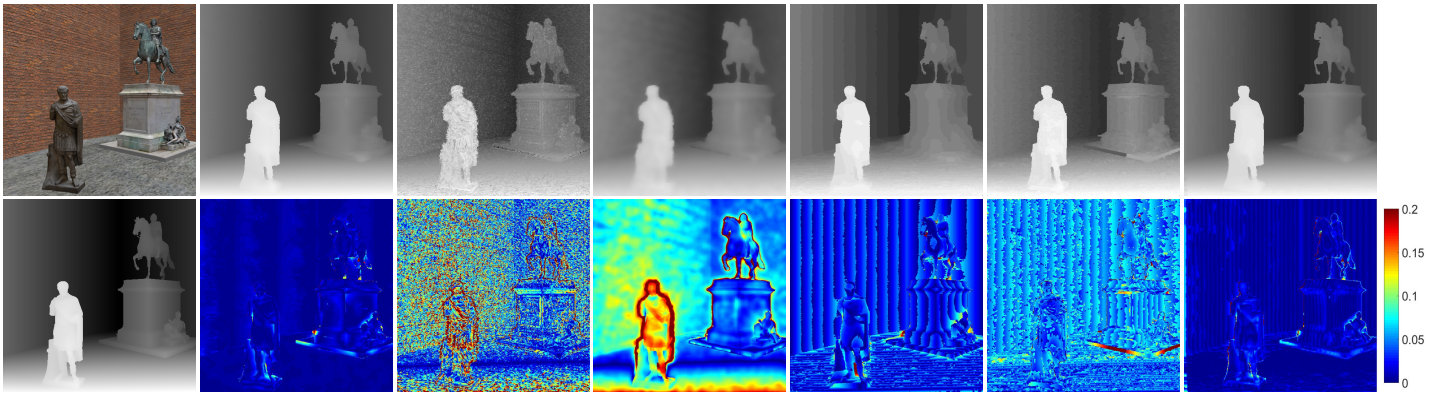
As shown in Fig. 2, for each light field image pair captured from a real scene, we show the central subaperture view of L_1, the initial coarse disparity maps from the structure tensor technique [15], the disparity maps estimated by the stereo matching method [32], and the final disparity maps optimized by our method. We also compare our method with existing methods, [15], [19] and [36]. Tab. II shows the corresponding calibrated parameters for each light field image pair.

Fig. 2a shows the central view of L_1 for each light field image pair. Fig. 2b shows the initial coarse disparity map for



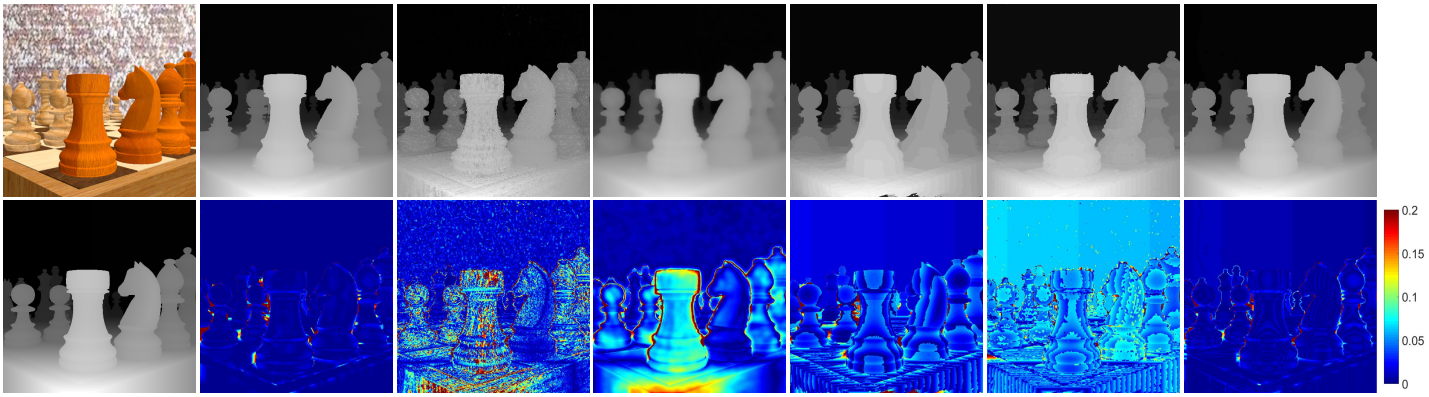
(a) Central views (b) Structure tensor (c) Tian [32] (d) Wanner [15] (e) Jeon [19] (f) Wang [36] (g) Our method

Fig. 2: Disparity map estimation for light field image pairs captured in real scenes: (a) the central subaperture views in L_1; (b) the disparity maps computed by the structure tensor technique [15]; (c) stereo matching results from [32]; (d) results from [15]; (e) results from [19]; (f) results from [36]; (g) disparity maps from our method.



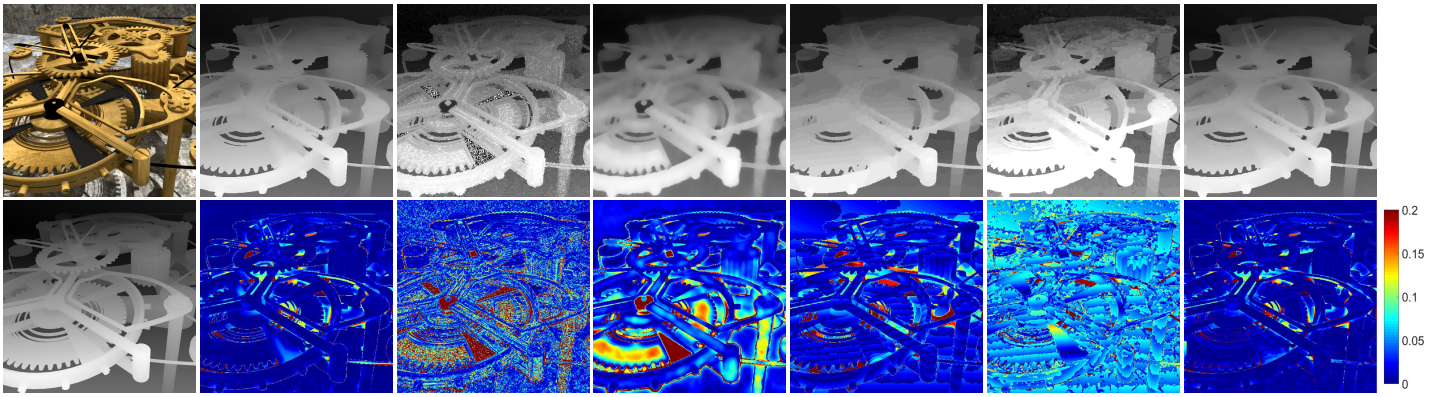
(a) Central view (b) Stereo matching (c) Structure tensor (d) Wanner [15] (e) Jeon [19] (f) Wang [36] (g) Our method

Fig. 3: Quantitative evaluation of the disparity estimation for our method.



(a) Central view (b) Stereo matching (c) Structure tensor (d) Wanner [15] (e) Jeon [19] (f) Wang [36] (g) Our method

Fig. 4: Quantitative evaluation of the disparity estimation for our method.



(a) Central view (b) Stereo matching (c) Structure tensor (d) Wanner [15] (e) Jeon [19] (f) Wang [36] (g) Our method

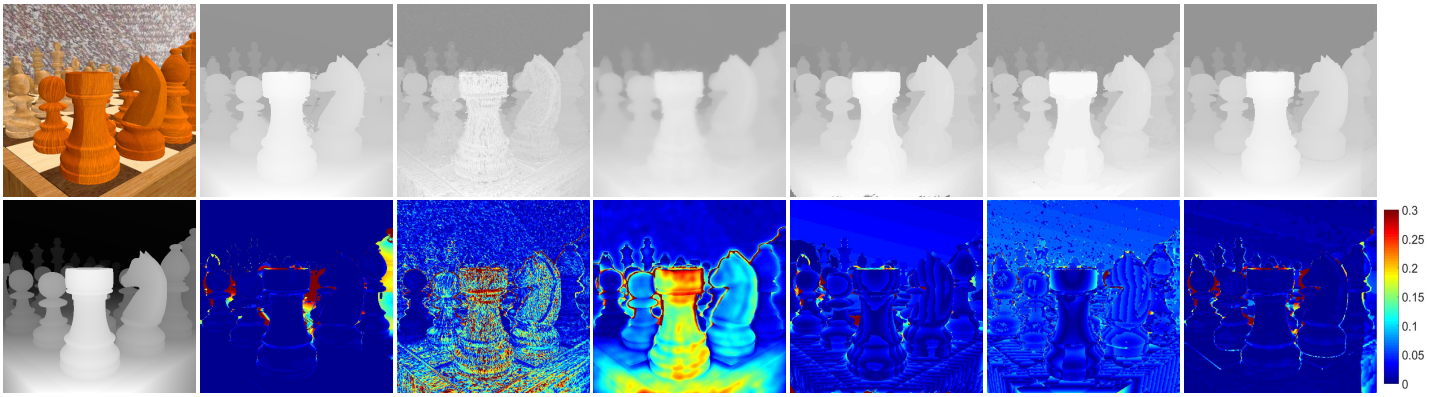
Fig. 5: Quantitative evaluation of the disparity estimation for our method.

Fig. 2a by adopting the structure tensor technique [15]. This type of disparity map always contains a substantial amount of noise, especially in the homogeneous regions. Fig. 2c shows the disparity map produced by the stereo matching method [32] for the two central subaperture views of L_1 and L_2. The disparity map has been mapped to the light field disparity range by adopting Eq. 2. This type of disparity map has smooth and reliable disparity values in the homogeneous regions. Therefore, there is less noise in these disparity maps. However, the comparison of these disparity maps with the disparity map in Fig. 2b reveals that this type of disparity map contains few details and a low depth contrast for objects in different depth layers. Fig. 2d, 2e and 2f show the disparity maps separately produced by the different methods: [15], [19] and [36]. Fig. 2g shows the disparity map obtained from our method. Our disparity maps are more precise, with less noise and a higher depth contrast than those produced by the other methods.

We also quantitatively evaluate our disparity estimation method on light field image pairs generated from virtual scenes, as shown in Figs. 3, 4, 5, 6 and 7. Taking Fig. 3 as an example, Fig. 3a shows the central subaperture view of L_1 and its ground truth disparity map. Fig. 3b shows the disparity map estimated by a stereo matching method [32] and the residual map, which represents the absolute value of the difference between the disparity map and the ground truth disparity map. Fig. 3c shows the disparity map extracted from EPIs by adopting the structure tensor technique [15] and the residual map generated in the same way as in Fig. 3b. Similarly, we also show the disparity maps estimated by the existing methods [15], [19], [36] and our method, and the corresponding residual maps generated as above in Figs. 3d, 3e, 3f and 3g, respectively.

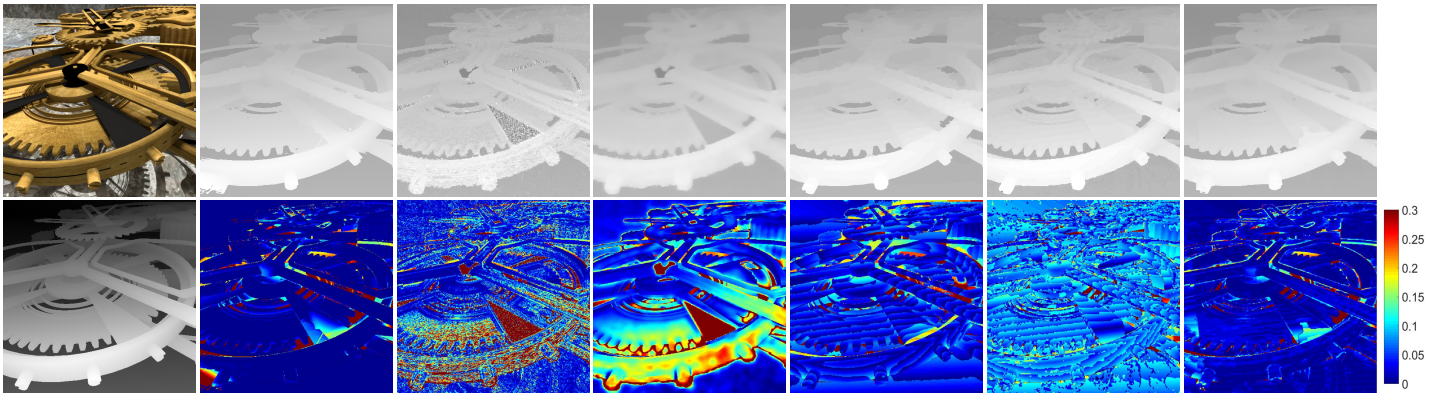
The stereo matching method [32] and our method outperform all other methods. In contrast with the disparity maps estimated from the real scene light field image pairs (Fig. 3c), the stereo matching method [32] performs very well on the synthetic light field image pairs. Although the overall performances of our method and [32] are similar in the Figs. 3, 4 and 5, the disparity maps obtained by our method have few obvious errors (in bright red) in near regions. In addition, our method clearly outperforms the stereo matching method in Figs. 6 and 7, especially in the object boundaries with abrupt disparity changes.

These two sets of experimental results separately tested on light field images from real and virtual scenes demonstrate that our method outperforms the state-of-the-art methods, [15], [19] and [36]. Therefore, our method can provide the subsequent stereoscopic image synthesis with accurate disparity maps.



(a) Central view (b) Stereo matching (c) Structure tensor (d) Wanner [15] (e) Jeon [19] (f) Wang [36] (g) Our method

Fig. 6: Quantitative evaluation of the disparity estimation for our method.



(a) Central view (b) Stereo matching (c) Structure tensor (d) Wanner [15] (e) Jeon [19] (f) Wang [36] (g) Our method

Fig. 7: Quantitative evaluation of the disparity estimation for our method.

NO.	Parameters for each LFI pair from the real scene					
	f(pixel)	B(mm)	b(mm)	d_s(pixel)	d_{min}	d_{max}
1	1535	20	1.0644	0.103552	-0.1196	0.5048
2	1189	20	0.8562	0.104805	-0.1669	0.7426
3	987	30	1.6507	0	-0.0346	0.8
4	987	20	1.1182	0	-0.3	0.8
5	987	20	1.2425	0	-0.3	0.8
6	1258	10	0.8801	0.005	-0.3	0.8
7	808	30	1.6569	0	-0.3	0.8
8	1518	5	0.8479	3.3111	-3.3111	1.2

TABLE II: Calibrated camera parameters and disparity range for each light field image (LFI) pair shown in Fig. 2.

NO.	Parameters for each LFI pair from the virtual scene					
	f(pixel)	B(mm)	b(mm)	d_s(pixel)	d_{min}	d_{max}
Fig. 3	585.1	50	10	0.5851	0.2386	1.6094
Fig. 4	658.3	100	15	0.4937	-0.2447	1.6930
Fig. 5	658.3	200	25	0.4114	-0.1397	0.8270
Fig. 6	585.1	30	1	0.0585	0.2196	1.8131
Fig. 7	512	30	1	0.0512	0.1115	1.9759

TABLE III: Camera parameters and disparity range for each light field image (LFI) pair from the virtual scene.

1) Ablation Study: We have conducted an ablation study to evaluate the necessity and effectiveness of our proposed coherence energy terms (the Eqn. 17 and Eqn. 19). The experimental results are shown in Figs. 8- 11.

Figs. 8 and 9 show the experimental results for light field images from real scenes. Taking Fig. 8 as an example, the first row shows disparity maps produced by our method taking both two coherence constraints. The disparity maps exhibited in this row are very consistent for different subaperture views. In contrast, the second row shows disparity maps produced by our method with only the coherency term defined for two neighboring subaperture views of the same light field image (Eqn. 17). And the third row shows the disparity maps produced by our method with only the coherence constraint defined for corresponding subaperture views separately within the two light field images (Eqn. 19). The disparity maps shown in the second row are similar to the disparity maps exhibited in the first row, which means that the coherence constraint for neighboring subaperture views can make sure to produce accurate disparity maps for each single light field image. The disparity maps produced by our method without taking coherence constraints are shown in the fourth row. Subfigures of disparity maps in the third and fourth rows show that the disparity for different subaperture views are inconsistent while without taking the coherence constraint for neighboring subaperture views. The situations in Fig. 9 are similar, which demonstrate the necessity and effectiveness of the disparity coherence constraints of our method.

We also show experimental results for light field images of virtual scenes with ground truth disparity maps. Actually,

these images represent the absolute value of the difference between the estimated disparity map and the ground truth disparity map. In Fig. 10, the disparity maps produced by our method with both disparity coherence constraint shown in the first row and those optimized by our method with only the disparity coherence constraint for neighboring subaperture views are more accurate and consistent than those in the third and fourth rows. This example demonstrates that the disparity coherence constraint for neighboring subaperture views are more important for accuracy and coherence of disparity maps for subaperture views within the same light field images. The situations in Fig. 11 are similar to Fig. 10.

B. Stereoscopic Image Generation Results

Based on the above optimized disparity maps for every input light field image pair, our method generates the desired stereoscopic images that satisfy the desired convergence requirement (Figs. 12 and 15-13).

In Fig. 12a, we show the stereoscopic images consisting of the two central subaperture views of each input light field image pair. These stereoscopic images can be seen as the reference stereoscopic images. There are no negative disparity values for these stereoscopic images, which means that the 3D scene conveyed by these stereoscopic images appears completely in front of the screen. Figs. 12b, 12c and 12d show the stereoscopic images generated by our method with various convergence requirements. The corresponding parameters for generating these stereoscopic images are shown in Tab. IV. By adjusting the convergence for the virtual stereo viewpoint pair, our method can flexibly control the target disparity range. Since the rotation angle for the optical axes of the virtual stereo camera pair for each generated stereoscopic image is very small, it is difficult for the photographers to manually control the rotation of the light field camera pair in the capture stage. Therefore, our method allows users to flexibly adjust the convergence for the target stereoscopic images, which is similar to what photographers do in the capture stage.

To quantitatively evaluate our method for stereoscopic image generation with convergence control ability, we generate a stereoscopic image from light field image pairs of the virtual scene (Figs. 13-15). Taking Fig. 13 as an example to demonstrate the effectiveness of our method, we perform not only the stereo viewpoint pair rotation (Fig. 13b) but also the stereo viewpoint pair shift and rotation (Fig. 13c). The parameter t (in mm) represents the shift in the two virtual viewpoints of the target stereoscopic images that separately move toward the center of the original stereo viewpoint pair before undergoing a rotation β (in degrees). While performing the viewpoint shift, there will be more disoccluded regions that need to be filled, which is a challenging problem for the view synthesis method. For each example, we show the stereoscopic images generated by our method without adopting view synthesis optimization, those generated by performing our view synthesis optimization. We compare the generated stereoscopic images with the ground truth stereoscopic images rendered by blender [35]. The PSNR values are listed in Tab. V. Our method with our own optimized disparity maps can produce much better stereoscopic images than those generated with the disparity maps obtained from [15].

In this set of experimental results, there is no obvious noise/distortion in the stereoscopic images generated by our method, which demonstrates the effectiveness of our method. Since our stereoscopic view synthesis method is based on the fast total variational optimization technique for light field image novel view synthesis [15], it produces the optimized stereoscopic images very efficiently.

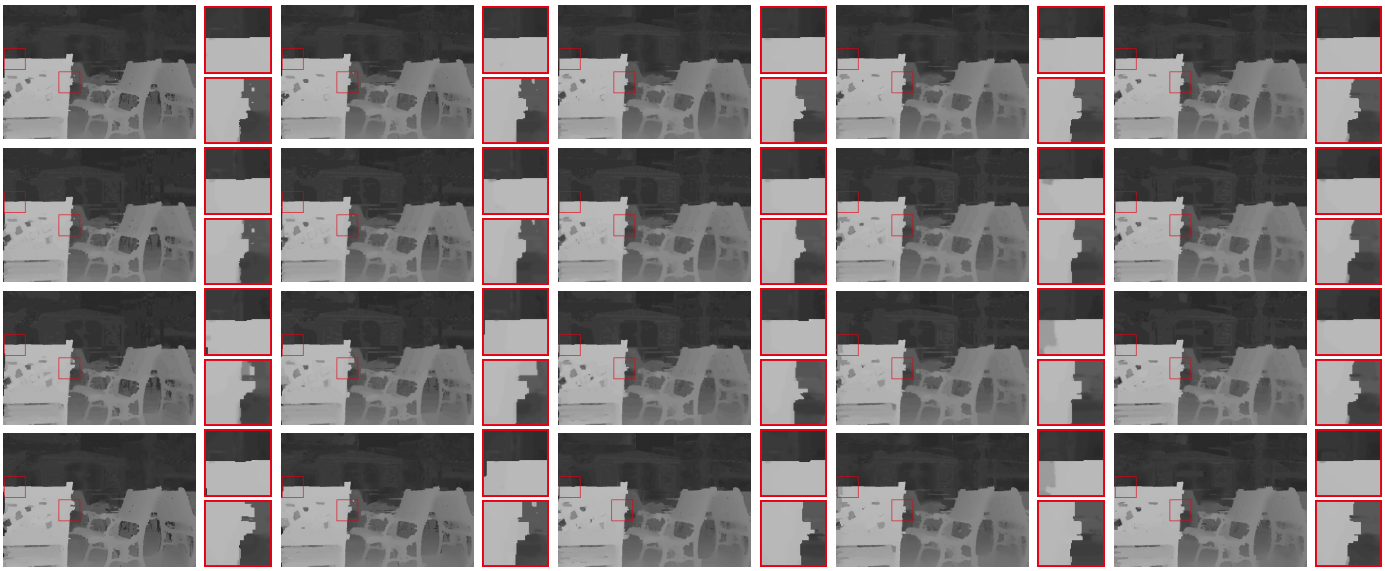
NO.	Fig. 12b		Fig. 12c		Fig. 12d	
	Z_c (m)	β (°)	Z_c (m)	β (°)	Z_c (m)	β (°)
1	6	0.095	3	0.1911	1.5	0.38
2	3	0.19	1.5	0.38	0.75	0.7643
3	12	0.0717	3	0.2866	1.5	0.5732
4	3	0.19	1.5	0.38	1	0.57
5	4	0.1433	2	0.2866	1	0.57
6	6	0.0478	1	0.2866	0.5	0.5732
7	12	0.0717	3	0.2866	1.5	0.5732
8	1	0.1433	0.5	0.2866	0.25	0.5732

TABLE IV: Parameters for the generated stereoscopic images shown in Fig. 12: Z_c is the focal distance for the stereo camera pair. $|\beta_1| = |\beta_2| = \beta$ means that left and right viewpoints for each desireded stereoscopic image take the same convergent rotation angle value but in opposite directions.

VII. Conclusion and Future Work

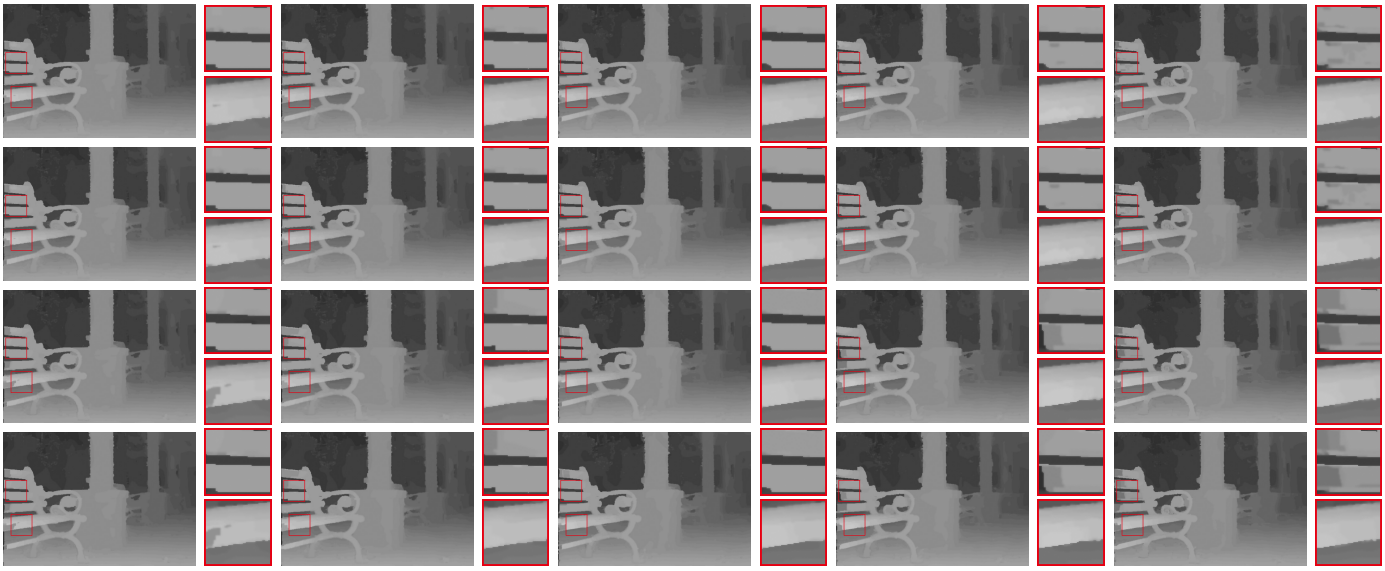
In this paper, we propose a method to generate the desired stereoscopic images with convergence control (i.e., these images would satisfy the specified convergence requirements) from a light field image pair. We first co-optimize the disparity maps for the input light field image pair and thereby obtain more accurate disparity maps. Based on these estimated disparity maps, our method then generates the target stereoscopic images that satisfy the desired convergence requirements. Extensive experiments have been conducted on real- and virtual-scene light field images to demonstrate the effectiveness of our method.

In future work, we will aim to speed up our method, particularly the disparity estimation module, by making it run on a GPU card at an interactive rate. Moreover, we will create a super-resolution function for our stereoscopic image generation method.



(a) subview [4,2] (b) (c) subview [4,3] (d) (e) subview [4,4] (f) (g) subview [4,5] (h) (i) subview [4,6] (j)

Fig. 8: Experimental results for ablation study: First row shows disparity maps produced by our disparity estimation method. Second row shows disparity maps optimized by our method with only the coherence constraint for two neighboring subaperture views within the same light field image (Eqn. 17), and the third row shows disparity maps produced by our method with taking only the coherence constraint for corresponding subaperture views separately within the two light field images (Eqn. 19). Fourth row shows disparity maps produced by our method taking none of the two coherence constraint. (b), (d), (f) and (h) are subfigures of (a),(c),(e) and (g), respectively. All these disparity maps are from the subaperture views of L_1.



(a) subview [4,2] (b) (c) subview [4,3] (d) (e) subview [4,4] (f) (g) subview [4,5] (h) (i) subview [4,6] (j)

Fig. 9: Experimental results for ablation study. The layout and description of images in this example are similar to those in Fig. 8.

Approach: Our method	Evaluation (PSNR)									
	Fig. 13b	Fig. 13c	Fig. 14b	Fig. 14c	Fig. 15b	Fig. 15c	Fig. 16b	Fig. 16c	Fig. 17b	Fig. 17c
without view synthesis optimization	26.5	25.2	34.1	25.10	31.27	30.66	28.8	24.83	38.7	30.05
utilizing disparity estimated by [15]	36.4	32.7	39.9	35.95	39.12	37.32	40.8	37.93	38.7	37.2
using our optimized disparity maps	36.4	**33.6**	**40.0**	**37.14**	39.12	**37.66**	40.8	**38.92**	38.7	**37.75**

TABLE V: Quantitative evaluation of our method for stereoscopic images generated from light field image pairs of the virtual scenes.

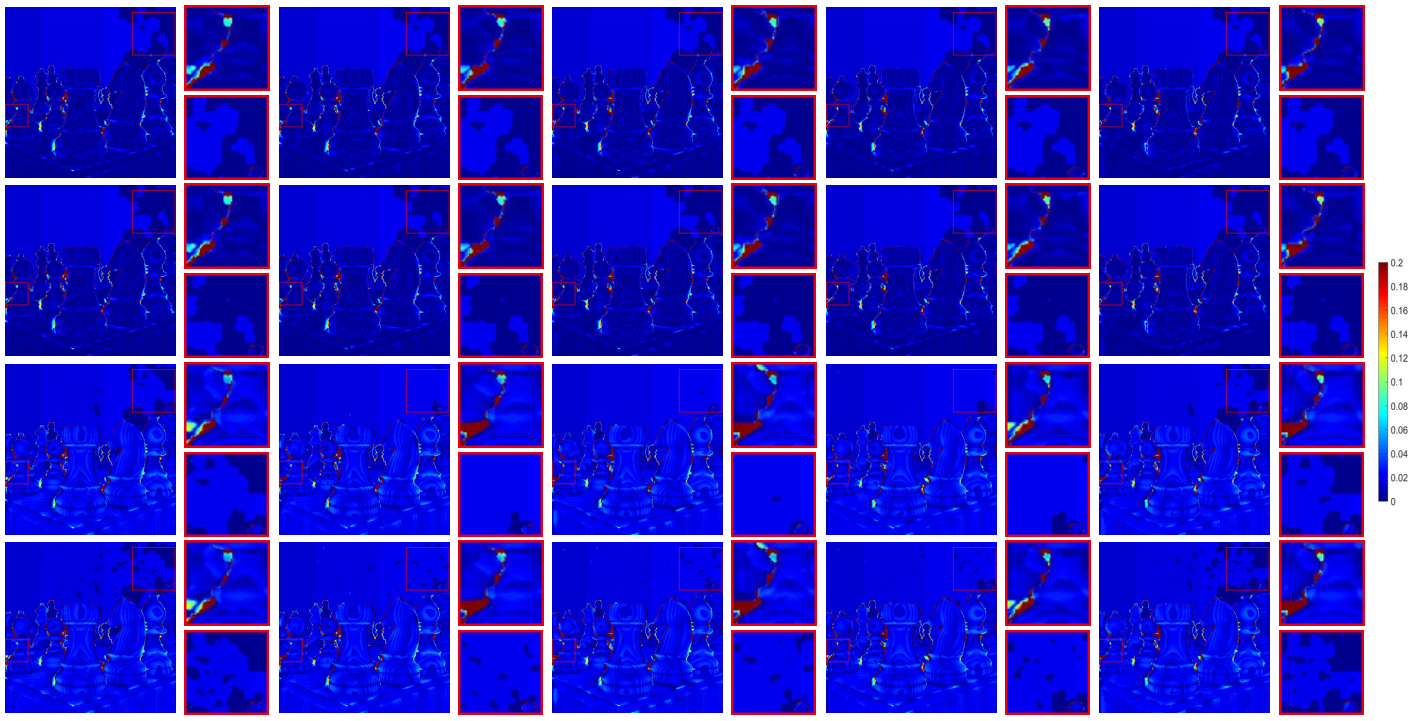


Fig. 10: Experimental results for ablation study. The layout and description of images in this example are similar to those in Fig. 8. These images represent the absolute value of the difference between the estimated disparity map and the ground truth disparity map.

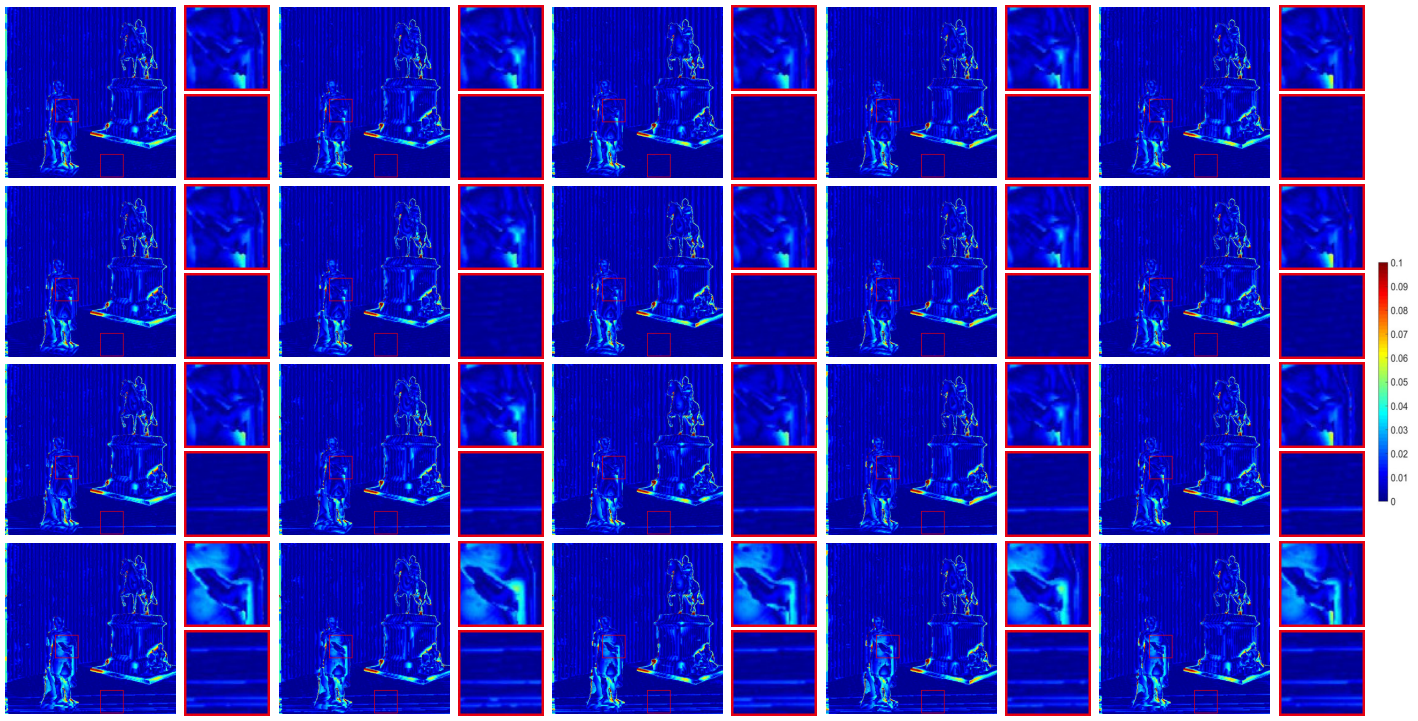
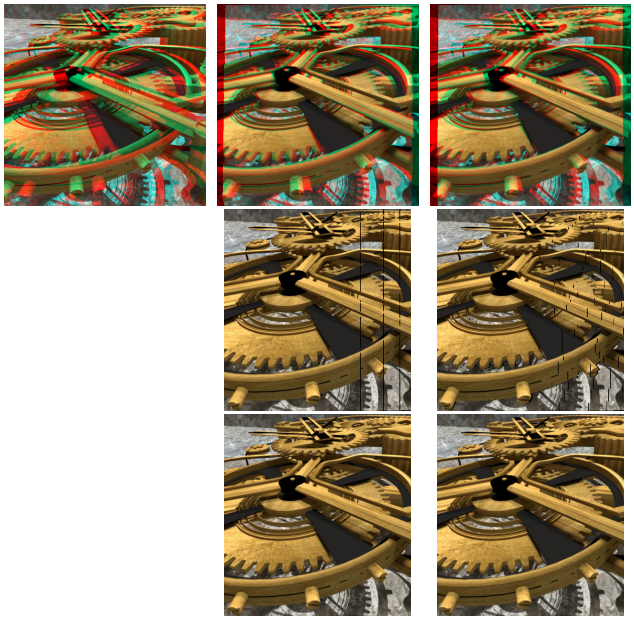


Fig. 11: Experimental results for ablation study. The layout and description of images in this example are similar to those in Fig. 8. These images represent the absolute value of the difference between the estimated disparity map and the ground truth disparity map.

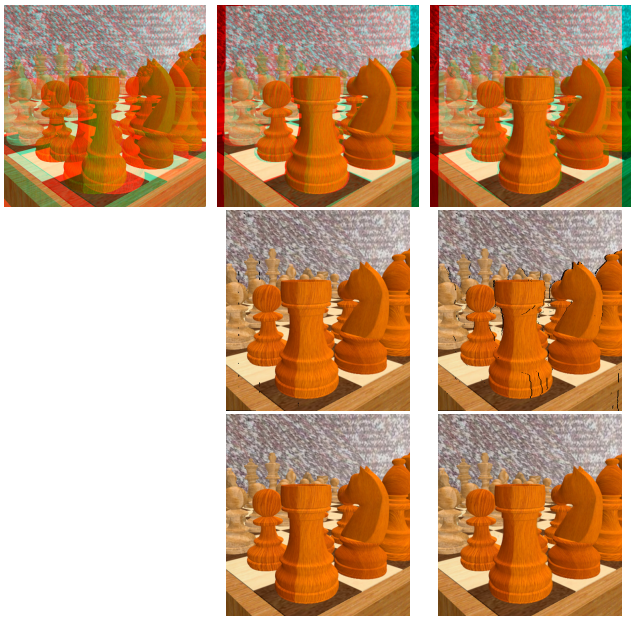
(a) (b) (c) (d)

Fig. 12: Stereoscopic image generated by our method with various convergence settings. (a) shows the original stereoscopic images. (b) (c) and (d) show the stereoscopic images generated by our proposed method with various convergence parameters as listed in Tab. IV.



(a) Original (b) $t = 0$, $\beta = 2$ (c) $t = 3$, $\beta = 2$

Fig. 13: Stereoscopic images generated by our method. (a) shows the original stereoscopic images. (b) and (c) show our produced stereoscopic images with different shift and convergence settings. From top to bottom: the first row shows the stereoscopic images; the second row shows the warped left views before our view synthesis optimization; our optimized left views are shown in the third row. t is in mm and β is in degrees.



(a) Original (b) $t = 0$, $\beta = 2$ (c) $t = 3$, $\beta = 2$

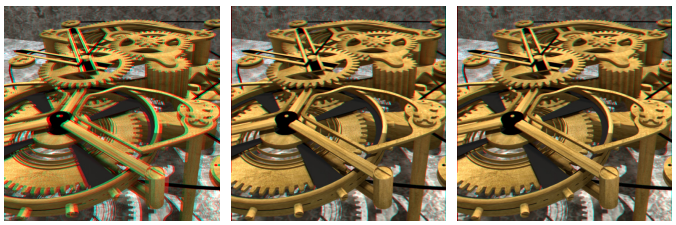
Fig. 14: Stereoscopic images generated by our method. The layout and description of images in this example are similar to those in Fig. 13, except that the second and third rows show the corresponding right views before and after our optimization.

(a) Original (b) $t = 0$, $\beta = 0.3$ (c) $t = 10$, $\beta = 0.3$

Fig. 15: Stereoscopic images generated by our method with various convergence settings. t is in mm and β is in degrees.

(a) Original (b) $t = 0$, $\beta = 0.5$ (c) $t = -20$, $\beta = 0.5$

Fig. 16: Stereoscopic images generated by our method with various convergence settings. t in mm and β is in degrees.



(a) Original (b) $t = 0$, $\beta = 0.23$ (c) $t = 40$, $\beta = 0.23$

Fig. 17: Stereoscopic images generated by our method with various convergence settings. t is in mm and β is in degrees.

REFERENCES

[1] "Lytro." [Online]. Available: https://en.wikipedia.org/wiki/Lytro

[2] "Raytrix." [Online]. Available: https://raytrix.de/

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

[4] 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, Oct 2017.

[5] P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, and H.-P. Seidel, "A perceptual model for disparity," *ACM TOG*, vol. 30, no. 4, 2011.

[6] T. Yan, R. Lau, Y. Xu, and L. Huang, "Depth mapping for stereoscopic videos," *IJCV*, vol. 102, no. 1-3, pp. 293–307, Mar. 2013.

[7] 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, June 2011.

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

[9] S. Heinzle, P. Greisen, D. Gallup, C. Chen, D. Saner, A. Smolic, A. Burg, W. Matusik, and M. Gross, "Computational stereo camera system with programmable control loop," *ACM TOG*, vol. 30, no. 4, 2011.

[10] S. Koppal, C. L. Zitnick, M. Cohen, S. B. Kang, B. Ressler, and A. Colburn, "A viewer-centric editor for 3d movies," *IEEE CG&A*, vol. 31, no. 1, pp. 20–35, 2011.

[11] A. Smolic, P. Kauff, S. Knorr, A. Hornung, M. Kunter, M. Muller, and M. Lang, "Three-dimensional video postproduction and processing," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 607–625, 2011.

[12] C. Kim, A. Hornung, S. Heinzle, W. Matusik, and M. Gross, "Multi-perspective stereoscopy from light fields," *ACM TOG*, vol. 30, no. 6, Dec 2011.
[13] C. Kim, U. Mller, 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.
[14] 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.
[15] 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.
[16] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE TPAMI*, vol. 23, no. 11, pp. 1222–1239, 2001.
[17] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
[18] H. Lin, C. Chen, S. Bing Kang, and J. Yu, "Depth recovery from light field using focal stack symmetry," in *Proc. IEEE CVPR*, 2015, pp. 3451–3459.
[19] H. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y. W. Tai, and I. Kweon, "Accurate depth map estimation from a lenslet light field camera," in *Proc. IEEE CVPR*, June 2015, pp. 1547–1555.
[20] M. S. Sajjadi, R. Köhler, B. Schölkopf, and M. Hirsch, "Depth estimation through a generative model of light field synthesis," in *Proc. German Conference on Pattern Recognition*, 2016, pp. 426–438.
[21] C. Hazirbas, L. Leal-Taixé, and D. Cremers, "Deep depth from focus," *arXiv preprint arXiv:1704.01085*, 2017.
[22] S. Heber, W. Yu, and T. Pock, "Neural epi-volume networks for shape from light field," in *Proc. IEEE CVPR*, 2017, pp. 2252–2260.
[23] X. Guo, Z. Chen, S. Li, Y. Yang, and J. Yu, "Deep eyes: Binocular depth-from-focus on focal stack pairs," *arXiv:1711.10729*, 2017.
[24] Y. Yoon, H. Jeon, D. Yoo, J. Lee, and I. Kweon, "Learning a deep convolutional network for light-field image super-resolution," in *Proc. IEEE ICCV Workshop*, Dec 2015, pp. 57–65.
[25] 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.
[26] 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*, June 2016, pp. 5515–5524.
[27] N. K. Kalantari, T.-C. Wang, and R. Ramamoorthi, "Learning-based view synthesis for light field cameras," *ACM TOG*, vol. 35, no. 6, Nov 2016.
[28] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: exploring photo collections in 3d," in *ACM TOG*, vol. 25, no. 3, 2006, pp. 835–846.
[29] N. Snavely, R. Garg, S. M. Seitz, and R. Szeliski, "Finding paths through the world's photos," in *ACM TOG*, vol. 27, no. 3, 2008, p. 15.
[30] M. Lambooij, M. Fortuin, I. Heynderickx, and W. IJsselsteijn, "Visual discomfort and visual fatigue of stereoscopic displays: A review," *Journal of Imaging Science and Technology*, vol. 53, no. 3, pp. 30 201–1, 2009.
[31] M. Lambooij, W. A. IJsselsteijn, and I. Heynderickx, "Visual discomfort of 3d tv: Assessment methods and modeling," *Displays*, vol. 32, no. 4, pp. 209–218, 2011.
[32] T. Taniai, Y. Matsushita, Y. Sato, and T. Naemura, "Continuous 3d label stereo matching using local expansion moves," *IEEE TPAMI*, 2017, (accepted).
[33] A. Hosni, M. Bleyer, C. Rhemann, M. Gelautz, and C. Rother, "Real-time local stereo matching using guided image filtering," in *Proc. IEEE ICME*, 2011, pp. 1–6.
[34] C. Hahne, A. Aggoun, V. Velisavljevic, S. Fiebig, and M. Pesch, "Baseline and triangulation geometry in a standard plenoptic camera," *IJCV*, Aug 2017.
[35] "Blender." [Online]. Available: https://www.blender.org/
[36] T. Wang, A. A. Efros, and R. Ramamoorthi, "Occlusion-aware depth estimation using light-field cameras," in *Proc. IEEE ICCV*, 2015, pp. 3487–3495.

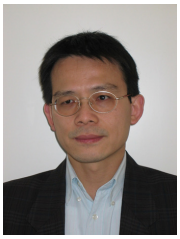
Tao Yan received his Ph.D. degree in computer science from City University of Hongkong (CityU) and University of Science and Technology of China (USTC), in 2013. He is now a lecturer at Jiangnan University in China. His research interests include image process, computer vision and machine learning.

Yiming Mao received his bachelor degree in digital media technology from Changshu Institute of Technology, China, in 2016. He is now pursuing his MS degree at Jiangnan University. His research interests include computer vision and image processing.

Jianming Wang received his bachelor degree in digital media technology from Jiangnan University, China, in 2016. He is now pursuing his MS degree at Jiangnan University. His research interests include computer vision and image processing.

WenxiLiu is currently an Associate Professor at the College of Mathematics and Computer Science, Fuzhou University, China. He earned his Ph.D. degree at City University of Hong Kong. His research interests include visual tracking, crowd analysis, and crowd simulation.

Xiaohua Qian received his Ph.D. degree in Electronic Engineering from Jilin University, Changchun, China, in 2012. He has been an Associated Professor at the BME at Shanghai Jiao Tong University since 2018. Before joining SJTU, Dr. Qian worked at Wake Forest University's School of Medicine as a Research Fellow and then worked at the University of Texas Health Science Center at Houston as an Assistant Professor. His primary research interests and areas of expertise are image processing and machine learning.



Rynson W.H. Lau received his Ph.D. degree from the University of Cambridge. He was on the faculty of Durham University and is now with City University of Hong Kong.

Rynson serves on the Editorial Boards of Computer Graphics Forum, and Computer Animation and Virtual Worlds. He has served as the Guest Editor of a number of journal special issues, including ACM Trans. on Internet Technology, IEEE Trans. on Multimedia, IEEE Trans. on Visualization and Computer Graphics, and IEEE Computer Graphics & Applications. Rynson's research interests include computer graphics, image processing and computer vision.